THE UNIVERSITY OF ALBERTA

A RELIABILITY STUDY OF HIGH SCHOOL ESSAY SCORING

BY

ADELL MONA NYBERG

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF SECONDARY EDUCATION

EDMONTON, ALBERTA
FALL, 1987

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-41025-6

THE UNIVERSITY OF ALBERTA

RELEASE FORM

©

NAME OF AUTHOR: ADELL MONA NYBERG

TITLE OF THESIS: A RELIABILITY STUDY OF HIGH SCHOOL ESSAY SCORING

DEGREE: DOCTOR OF PHILOSOPHY

YEAR THIS DEGREE GRANTED: 1987

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Adell M. Nyberg

16516 - 78 Avenue
Edmonton, Alberta
Canada
T5R 3E6

Date: October 5, 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled A RELIABILITY STUDY OF HIGH SCHOOL ESSAY SCORING submitted by ADELL MONA NYBERG in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

John E. Oster
(Supervisor)

Douglas V. Parker

Jim Parsons

Joyce Edwards

(External Examiner)

Date: Oct. 5, 1987

ABSTRACT

This study had two main purposes. The first was to determine the factors in each of four areas that tended to reduce reliability of scoring of grade twelve essay examinations. These areas were compositions, readers, scoring variables, and writing tasks. The second purpose was to determine the reliability of the essay examination as a whole.

The study utilized three samples of papers. One consisted of seven papers scored by seventy-five markers, and the second of ten essays scored by sixty-four markers from the January, and June, 1986, scoring sessions. These samples were used to study four sources of unreliability of scoring. The third sample of ninety-six compositions was used to determine the reliability of the test as a whole, and to determine the improvement in reliability that resulted from having compositions read more than once.

The major problem related to the compositions was a halo effect among the readers. This effect was caused by numerous mechanical errors, approval or disapproval of the views expressed by the writer, unusual compositions, sentimental or emotional content, essay length, false indicators of complex thought, and expression typical of English as second language students.

With respect to reader-based problems, it was found that some readers tended not to agree with the general consensus; that is, their correlations with other readers were low.

Other readers were consistently reliable in their scoring. Also, some readers tended to award marks that were higher or lower than those of other readers.

Scoring variables associated with the more complex writing tasks (organization, thought and detail) were scored less reliably than other variables.

Of the three writing tasks, functional writing was scored least reliably, and response to visual communication most reliably.

Reliability of the essay examinations was estimated through use of coefficient alpha. A modest increase from 0.796 to 0.816 was recorded from the first reading to the final reading; however, it was evident that the first value was inflated by the halo effect. Strong evidence of the value of scoring the essays more than once was presented.

Recommendations were made with a view to improving reliability of scoring.

ACKNOWLEDGEMENTS

The author would like to recognize the assistance of the Student Evaluation and Records Branch of Alberta Education in offering for analysis compositions written by English 33 students. The author is especially indebted to Mr. Thomas Dunn and to Dr. David Wasserman.

The author would also like to acknowledge the support of her supervisor, Dr. John Oster, who gave valuable assistance in the planning and revision stages of this study, and to the members of her examining committee who were both positive and helpful.

Special thanks go to the author's husband, Richard Ferchoff, to her in-laws, Violet and Henry Ferchoff, to her great aunt, Gertrude Baldwin, and to her parents, Mona and Verner Nyberg. Her father's tutoring in the area of educational measurement helped to make the entire procedure enjoyable; to him is owed a great debt of gratitude.

TABLE OF CONTENTS

CHAPTER PAGE

I. THE PROBLEM 1

Background and Significance 1

Statement of the Problem 3

A. Compositions Having Low Scorer Reliability. 5

B. Assessing Test Reliability 6

Assumptions 6

Limitations 7

Definitions of Terms 9

A. Holistic Scoring 9

B. Focused Holistic Scoring 10

C. Analytic Scoring 10

D. Atomistic Scoring 10

E. Scoring Descriptors 11

F. English 30 and 33 11

G. Halo Effect 12

Summary 12

II. REVIEW OF THE LITERATURE 13

Introduction 13

Reliability Studies 14

Summary of Literature and Extent of Application by Alberta Education 36

III. DESIGN OF THE STUDY 42

General 42

Sampling 42

Data Gathered 44

CHAPTER PAGE

Reliability Review Procedures 46

Details of the Design 48

I. Compositions Having Low Scorer Reliability 51

A. Problems Related to the Compositions 51

B. Reader-Based Problems 54

C. Problems Related to the Scoring Variables 55

D. Problems Related to the Writing Tasks 55

II. Assessing Test Reliability 56

Summary 57

IV. OBSERVATIONS AND DISCUSSION OF RESULTS 58

I. Compositions Having Low Scorer Reliability 58

A. Problems Related to the Compositions . . 58

1. January, 1986, Reliability Review Papers 58

Paper Causing Greatest Problem - Jan. R.R. Paper #5 59

Paper Causing Second Greatest Problem - Jan. R.R. Paper #2 63

Paper Causing Third Greatest Problem - Jan. R.R. Paper #4 66

Paper Causing Fourth Greatest Problem - Jan. R.R. Paper #7 70

Paper Causing Fifth Greatest Problem - Jan. R.R. Paper #1 73

Paper Causing Sixth Greatest Problem - Jan. R.R. Paper #6 77

Paper Causing Seventh Greatest Problem - Jan. R.R. Paper #3 80

2. June, 1986, Reliability Review Papers 83

CHAPTER PAGE

Paper Causing Greatest Problem - June R.R. Paper #9 83

Paper Causing Second Greatest Problem - June R.R. Paper #6 88

Paper Causing Third Greatest Problem - June R.R. Paper #2 91

Paper Causing Fourth Greatest Problem - June R.R. Paper #8 94

Paper Causing Fifth Greatest Problem - June R.R. Paper #7 96

Paper Causing Sixth Greatest Problem - June R.R. Paper #1 98

Paper Causing Seventh Greatest Problem - June R.R. Paper #4 102

Paper Causing Eighth Greatest Problem - June R.R. Paper #5 104

Paper Causing Ninth Greatest Problem - June R.R. Paper #3 106

Paper Causing Tenth Greatest Problem - June R.R. Paper #10 108

3. Generalizations Regarding Problem Papers 109

B. Reader-Based Problems 112

1. Lack of Agreement With Other Readers 113

2. Easy or Severe Markers 119

3. Discrepant Spreads of Scores 120

4. Overall Results 123

C. Problems Related to the Scoring Variables 126

D. Problems Related to the Writing Tasks . 128

II. Assessing Test Reliability 132

Summary 135

V. CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS 138

CHAPTER PAGE

I. Compositions Having Low Scorer Reliability 138

A. Problems Related to the Compositions . . 138

Recommendation One 139

B. Reader-Based Problems 140

Recommendation Two 142

Recommendation Three 143

C. Problems Related to the Scoring Variables 143

Recommendation Four 144

Recommendation Five 144

D. Problems Related to the Writing Tasks . 144

Recommendation Six 146

II. Assessing Test Reliability 147

Recommendation Seven 149

Recommendation Eight 149

III. Implications 149

IV. Recommendations for Further Research . . . 151

Summary 153

BIBLIOGRAPHY 155

APPENDIX A. SCORING DESCRIPTORS--ENGLISH 33 161

APPENDIX B. RELIABILITY REVIEW PAPERS - JANUARY, AND JUNE, 1986 169

APPENDIX C. GRADE 12 DIPLOMA EXAMINATION - ENGLISH 33 PART A: WRITTEN EXPRESSION, JANUARY, AND JUNE, 1986 285

APPENDIX D. Table D-1 MEAN CORRELATIONS BY WRITING TASK FOR ALL READERS (JANUARY, 1986) Table D-2 MEAN CORRELATIONS BY WRITING TASK FOR ALL READERS (JUNE, 1986) 309

VITA 314

LIST OF TABLES

Table		Page
II-1	Summary of Results of Reliability Studies	37
II-2	Recommended Procedures Adopted by Alberta Education	39
II-3	Recommended Procedures Not Adopted by Alberta Education	41
III-1	Collection of Reliability Review Data	45
III-2	English 33 Diploma Examination Blueprint	49
III-3	Weighted Scores By Reader and Paper	50
III-4	Correlations Between Total Scores for All Pairs of Readers	52
III-5	Correlations Between Readers for All Scoring Variables	53
IV-1	Means and Standard Deviations of Scores Assigned to Reliability Review Papers	59
IV-2	Mean Correlations, Means, and Standard Deviations for All Readers (January,1986)	114
IV-3	Mean Correlations, Means, and Standard Deviations for All Readers (June, 1986)	116
IV-4	Distribution of Correlations by Category	118
IV-5	Distribution of Discrepancies in Means by Category	120
IV-6	Problem Readers (January, 1986)	124
IV-7	Problem Readers (June, 1986)	125
IV-8	Mean, Range, and Rank Order of Correlations by Scoring Variable (January, and June, 1986)	127
IV-9	Means of Correlations Between Readers by Writing Task	129
IV-10	Number of Times Readers' Mean Correlalations Appeared in Top or Bottom Ten	132

Table		Page
IV-11	Absolute Differences Between Initial Scores and Final Grades	135
D-1	Mean Correlations by Writing Task for All Readers January, 1986 (N = 75)	310
D-2	Mean Correlations by Writing Task for All Readers June, 1986 (N = 64)	312

LIST OF FIGURES

Figure / Page

IV-1 Standard Deviations for January, 1986, Scoring Session 121

IV-2 Standard Deviations for June, 1986, Scoring Session 122

IV-3 Distribution of Mean Correlations for Readers Over Writing Tasks (January, 1986) 130

IV-4 Distribution of Mean Correlations for Readers Over Writing Tasks (June, 1986) 131

CHAPTER I

THE PROBLEM

Background and Significance

Although examinations in oral form had been in use in China as early as 2200 B.C. when the emperor tested "his officials every third year to determine their fitness for continuing in office" (Dubois, 1970, p. 3), written tests were not introduced into European schools until the twelfth century at the time that paper began to replace papyrus as a recording medium (Dubois, 1970, p. 8). The use of written examinations for measuring student achievement in English language arts has been an acceptable practice for over a hundred and twenty-five years (Caldwell and Courtis, 1925, p. 37) and the utilization of written compositions to assess student achievement continues today.

With the use of written examinations comes the problem of the reliability of the scoring. This difficulty was noted as early as 1015 A.D. in China when examinations taken by public officials were graded by two readers while a third reader was required to "receive and reconcile the sealed grades" (Dubois, 1970, p. 4). The problem of the reliability of written compositions in the English language was identified as early as the 1880's (Englehart, 1950, p. 407) and it has been the topic of many research projects throughout the years. In general, each decade has produced at least one major reliability study (Edgeworth, 1890; Starch and Elliott, 1912; Darsie, 1922; Stalnaker, 1934; Anderson and Traxler, 1940; Finlayson, 1951; Diederich, French, and Carlton, 1961; Coffman, 1970; Sweedler-Brown, 1985), and the topic continues to be of importance to English educators today.

In recent years several provinces in Canada have returned to the utilization of government-sponsored examinations of the basic subject areas, and with these tests comes concern about the significance of the scoring methods. The province of Alberta began examining English language arts students at the high school leaving level in January, 1984. One half of the test that is administered by Alberta Education consists of an examination of written skills. When composition skills are tested on a large-scale basis, it is necessary for an examination authority to ensure that the papers are scored consistently. Alberta Education has attempted to do this by selecting only those markers who meet the criteria which have been established by the evaluation branch, by having the various school boards recommend outstanding candidates, and by choosing the most suitable people from the names which have been submitted. All markers are required to go through a training program, practice grading sample papers, and participate in consistency checking sessions. These procedures are carried out to improve the reliability of the scoring methods because the importance of treating students equitably is recognized.

Alberta's procedures related to the testing of writing ability involve thousands of students, and scores of markers and supervisors. The expenditures of money and human energy, plus the importance of the results to the careers of thousands of students, make periodic evaluations of the procedures imperative. This study assessed the scoring methods utilized by the province; however, it also had significance beyond the evaluative dimension. There were facets of the problem of the reliability of essay examinations that received little or no attention. Some of these facets were addressed in this research project.

The study dealt with essay examinations at the high school leaving level which were written by non-academic students, that is, students who were generally not college-bound. Compositions written by these individuals were chosen because researchers tended to concentrate on college- or university-bound students, and non-academic groups received little attention when the topic of inter-rater reliability was studied. The problem was probed by investigating some of the procedures employed by the province of Alberta in dealing with the difficulty of inter-rater reliability in a large-scale testing situation. The matter of the reliability of the essay test was also investigated. Data for the study were provided by Alberta Education based upon compositions and scores from the provincially administered English 33 examinations given in January, and June, 1986.

Statement of the Problem

This study focused upon two distinct problems. The first, and most important, was to identify features that posed grading problems which resulted in reduced reliability of scoring. In particular, four features that contributed to unreliability of scores were investigated--problem compositions, readers, scoring variables, and writing tasks.

Each aspect made a unique contribution to unreliability; however, no one of them was entirely independent of the others. Problem compositions as a source of unreliability included variation in grades caused by such things as idiosyncracies in expression, emotional overtones of the topic, and sex biases. Raters who ordinarily awarded highly similar grades might disagree on papers of a certain type. These types were referred to as problem papers in this study. Readers always disagree to some extent, even on straightforward compositions, and therefore constitute a source of unreliability of grading. Scoring variables also contribute to unreliability in their own way. Readers may agree highly on some variables and exhibit considerable disagreement on others. Finally, different compositions or writing tasks, produced by the same student, may vary in quality. Certain writing tasks may contribute more to scoring unreliability than others. These four sources were studied as they applied to papers selected to represent different scoring situations. A second problem of major importance was to determine the reliability of an essay examination as a whole, as opposed to limiting the study to the reliability of scoring. Reliability of scoring takes into account only the disagreement among raters. Reliability of the test as a whole involves the additional variation in scores caused by differences in quality of writing across variables being assessed. These problems were set in the context of a test of writing skills of a large number of non-academic candidates at the high school leaving level.

Specific objectives of the project are listed below. These are grouped according to the two problems mentioned, the matter of essays associated with low scorer reliability, and the matter of assessing test reliability.

A. Compositions Having Low Scorer Reliability

The first two specific objectives were related to the types of problems found in the papers.

1. Identify problem papers, that is, papers for which there was greater than average disagreement in grading.

2. Categorize the problem papers according to common characteristics that were related to low scorer reliability. The categories emerged as a result of the data produced.

The third objective related to the differences among readers as a source of unreliability.

3. Determine whether there were readers who tended to submit discrepant scores, and identify if the difficulty was general, or closely tied to the problem papers.

A fourth objective dealt with differences inherent in the scoring variables.

4. Assess the effectiveness with respect to reliability of each of a number of scoring variables that were commonly employed.

The fifth objective was concerned with differences in scores associated with different writing tasks.

5. Assess the effectiveness of each of several types of writing tasks that were employed in a written examination.

B. Assessing Test Reliability

The final two objectives were concerned with test reliability.

1. Calculate the reliability of a written test, involving eleven variables distributed over three topics, submitted by non-academic twelfth grade students.

2. Evaluate the improvement in reliability that resulted from combining scores submitted by three markers as opposed to scores submitted by one marker, only. The underlying question here was whether the improvement in reliability when changing from a single grading to three gradings was worth the additional expenditure of time and effort.

Assumptions

A number of assumptions were made at the outset of this study. The primary ones are stated below.

1. Aspects such as thought and detail, organization, matters of choice, and matters of convention are actually separate variables in writing which can be identified by teachers for assessment purposes.

2. The scoring variables specified and defined by Alberta Education constituted a valid measure of written

composition for non-academic grade twelve students in Alberta. It was assumed, for the purposes of this study, that these variables defined the field of composition at this level.

3. It was assumed that markers were making their best efforts to score papers according to the directions given by Alberta Education staff members.

4. The scoring procedures were intended, among other things, to improve the reliability of scoring over what would otherwise be achieved.

Limitations

Three samples of compositions were drawn for this study: one consisted of seven papers (the first review sample), the second consisted of ten papers (the second review sample), and the third consisted of ninety-six randomly selected papers (the reliability sample). A number of limitations were imposed on the study because of the nature of the samples.

With respect to the review samples, there were several important limitations. First, compositions written only by Alberta students were sampled for study. These essays were submitted by candidates at the high school leaving level (grade twelve) who were registered in a non-academic English course (English 33); that is, they were generally not college-bound students. Next, the sample compositions were written under rather specific conditions. Review samples were made up of compositions completed under provincial examination conditions during the January, and June, 1986, testing sessions. The written part of the examination is two-and-one-half hours in duration, involves three writing tasks, has no flexibility in choice of topic, allows only a dictionary and thesaurus as reference material, and is scored independently by readers who have taught the course at least twice and who are teaching English 33 during the time period from September, 1985, to June, 1986.

It follows that the results of the study related to the reliability sample may not be generalizable to compositions submitted in other countries, at other grade levels, by college-bound students, in examination conditions other than those described above, nor in compositions marked by people not involved in the teaching of the course related to the test.

In addition, it must be noted that the review samples were not selected through a random process (see pages 42 and 43). Instead, papers were chosen to represent the various responses made by students, including types of papers considered difficult to score as well as those perceived to be straightforward. The resulting limitation was that the study dealt with a restricted set of written responses. It is possible that some types were overlooked in the selection process and therefore were not studied. This limitation was not considered serious because of the experience of the people who made the initial selection of papers and the screening process that resulted in the final sample of papers which were employed during reliability review sessions.

Another limitation was associated with the scoring variables employed in the marking of the compositions. While different evaluation procedures had differing lists of variables, inevitably there was a high degree of overlap. The fact remained, however, that scores based upon different sets of variables differed (Diederich, 1974, p. 8), even with respect to the essay rankings that were produced. It must be noted, therefore, that the results of this study were generalizable only to a situation where a similar set of variables was employed. (The scoring descriptors are provided in Appendix A.)

Finally, with respect to limitations, it must be emphasized that this study was concerned with reliability of essay grades and essay examinations. It did NOT address the matters related to test validity.

Definitions of Terms

A. Holistic Scoring

Holistic scoring essentially consists of reading a document and assigning a grade on an impressionistic basis. Cooper (1977) states that with holistic scoring:

> The rater takes a piece of writing and either (1) matches it with another piece in a graded series of pieces, or (2) scores it for the prominence of certain features important to that kind of writing, or (3) assigns it a letter or number grade. The placing, scoring, or grading occurs quickly, impressionistically, after the rater has practiced the procedure with other raters (p. 3).

Holistic scoring is a popular method because it can be done quickly and is therefore less tedious than other methods (Coward, 1952, p. 84; Britton, 1966, p. 28).

B. Focused Holistic Scoring

Focused holistic scoring is similar to the holistic method; however, several scores are awarded representing different writing skills. Some of the commonly used aspects include mechanics, content, and style. Breland (1983) stated that with focused holistic scoring:

> The scoring might be done for each dimension after a single reading, or it might be done for each separately so as to minimize influences of one focus on the other. The number of focuses must of course be limited; otherwise, the procedure tends to be more like an analytical procedure (p. 9).

C. Analytic Scoring

The analytic method is much like focused holistic scoring, but the number of scoring aspects or dimensions is increased. Instead of two or three variables, the number may be ten or more, but each is scored on an impressionistic basis, usually on a scale from one to five. Cast (1939) describes the analytic method as having examiners "allot marks separately for each of the main aspects or elements of a good English composition, and to base their final marks on the total" (p. 260).

D. Atomistic Scoring

Atomistic scoring involves awarding or deducting points for specific details found in the composition. Fifty points may, for example, be blocked out for mechanics, then two points will be subtracted for each spelling error or punctuation error that is made. Additional points will be deducted for sentence errors and errors in form. The marks are totalled to obtain the final score. Coward (1952) summarizes atomistic scoring when she states:

> The reader's evaluation of a given composition is fragmentized; he is asked to make a number of specific and objective judgments on each composition, and these ratings are then summed to give the final grade for the paper (p. 81).

E. Scoring Descriptors

The scoring descriptors (also called scoring variables) are the categories outlined for markers so that grading of papers can take place according to pre-set guidelines. For this study the scoring descriptors have been outlined by Alberta Education in the January, and June, 1986, scorers' manuals. It is these scoring descriptors which establish the criteria for the marks that are awarded to the English 33 written examinations. The scoring descriptors for the three writing tasks can be found in Appendix A.

F. English 30 and 33

English 30 is a grade twelve or high school leaving course offered in the province of Alberta which is "more appropriate for students intending to pursue further academic studies at the university level" (Alberta Education, 1982, p. 6). English 33, another grade twelve high school leaving course, is "more appropriate for students intending to go to vocational school or to seek employment immediately after leaving high school" (Alberta Education, 1982, p. 6).

G. Halo Effect

The halo effect is a term applied when the rating of a characteristic is influenced by a previous rating given to another characteristic. Selltiz et al (1951), in describing the halo effect, state:

> If more than one characteristic of a person is to be judged, raters frequently carry over a generalized impression of the person from one rating to the next, or they try to make their ratings consistent. Thus, if a rater considers a person to be shy and he believes shy people to be poorly adjusted, he is likely to rate the person poorly adjusted as well as shy (p. 352).

Summary

Reliability of scoring of written compositions has been a matter of concern for centuries; however, studies in this area continue to be done. This study, using Alberta Education data related to English 33, a matriculation grade twelve high school leaving course, focused upon two broad problems. The first was to identify aspects that reduced the reliability of scoring. Difficulties related to papers, readers, scoring variables, and writing tasks were the four areas covered. The second problem was to determine the reliability of an essay examination as a whole, rather than limiting the study to the reliability of scoring.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

The effectiveness of written samples to assess the writing skills of individuals was summarized by Diederich (1974) when he stated:

> As a test of writing ability, no test is as convincing to teachers of English, to teachers in other departments, to prospective employers, and to the public as actual samples of each student's writing, especially if the writing is done under test conditions in which one can be sure that each sample is the student's own unaided work (p. 1).

A similar view was expressed by Britton, Martin, and Rosen (1966); however, they recommended the use of a reading test as well as a written examination to assess student abilities:

> . . .a candidate's ability to use his mother tongue in the written form is best tested by taking a sample of his performance in writing and in reading; that is to say that a composition test and a reading-comprehension test are the basic requirements and it ought not to be assumed that they need supplementing by grammatical or other forms of language test (p. 2).

Coffman (1971a) supported the other writers' opinions concerning the importance of task performance with his comment:

> The only way to assess the extent to which a student has mastered a field is to present him with questions or problems in the field and see how he performs. The scholar performs by speaking or writing. The essay examination constitutes a sample of scholarly performance; hence, it provides a direct measure of educational achievement (p. 273).

These types of statements help to demonstrate why the extended-answer examination has been used by educators for many years to assess the ability of students to write. Such statements also help to explain why, "despite more than a half a century of criticisms by educational measurement specialists, the essay remains a principal means of evaluation in courses of instruction of all types" (Breland, 1983, p. 7).

Because of the essay's significance in measuring the knowledge and abilities of students, the matter of the marking of compositions is of prime importance to educators. Problems associated with essay scoring have plagued both students and teachers since written examinations began to replace oral testing. Over the past seventy-five years, however, educators have become aware of the difficulties associated with the grading of written compositions and have carried out many studies in the attempt to eliminate or at least improve the situation.

Reliability Studies

One of the earliest studies that dealt with the problem of reliability was the work of Starch and Elliott conducted in 1912. These researchers had over three hundred markers score two final examinations in the subject area of English written by first year high school students. Scorers consisted of senior high school English teachers, English curriculum methods students, and students in an educational measurement class at the University of Chicago. Starch and Elliott found that ". . .the range of marks given by different teachers to the same paper may be as large as 35 or

40 points" (p. 454). They summarized their views regarding the problem of scoring written examinations by stating "such wide variations certainly impeach the reliability of the marks" (p. 456).

Typical studies of the nineteen twenties include one by Darsie (1922) and another by Hulten (1925). Both of these works involved the use of composition scales to provide guidelines for the scoring of papers. In Darsie's study, eighteen teachers college students marked compositions written by pupils in grades four to eight. Each scorer rated anywhere from ten to thirty papers and every paper was graded twice. Darsie concluded that the Willing Scale was "a very trustworthy instrument for measuring the 'story value' of compositions" (p. 89), and he reported a correlation of 0.884 for the two ratings of each paper. This high reliability might have been caused, in part, because the first grade appeared directly on the paper in view of the second marker.

Hulten used the Hudelson English Composition scale in his study. His results were not as favorable as those obtained by Darsie and he concluded that "teachers' marks are mere guesses, some good, some poor, some indifferent. Since they are mere guesses they are not sufficiently reliable to be used for promotion purposes" (p. 54).

A study by Stalnaker (1934) demonstrated that the author was aware of the problem of the reliability of scoring. He set out to improve reliability by having the scorers work in small groups because "it was hoped that agreement on standards would be reached more quickly and that greater reliability of reading would be achieved" (p. 219). These groups were, in effect, training the markers to become conscious of acceptable and unacceptable standards of writing for the compositions that were to be graded. Stalnaker reported a reliability coefficient of 0.88 after his markers had completed two readings of the same ninety papers.

Hartog and Rhodes (1936) examined a number of English tests written primarily at the high school leaving level. After examining the scores awarded to compositions by markers they stated:

> Some examiners marked consistently higher, some consistently lower, than the majority; others marked sometimes high, sometimes low. . . (p. 25).

In another part of their study Hartog and Rhodes had ten markers grade seventy-five papers using the impression method of scoring; the same markers then graded seventy-five additional papers using a detailed marking scheme. It was found that, in every case, the average of the impression grades was lower than that of the papers marked with the detailed marking guide. Based upon these findings, the researchers concluded:

> . . .the use of a detailed marking scheme does conduce to a closer approximation of the standards of examiners, but. . .it does nothing to reduce the element of random marking (p. 28).

Hartog and Rhodes noted that when four topic choices were given to students at the college entrance level, markers were consistently biased for or against particular topics that the pupils selected. They stated:

> The fate of a candidate in this type of examination is partly dependent on the particular examiner's reaction to the subject of the essay (p. 55).

This was the first study that documented the difficulty some graders have in scoring compositions written about subjects that they deem problematic.

A second study conducted by Stalnaker (1937) mentioned the importance of having students' names concealed on answer booklets and the significance of placing no marks on the compositions when they were being scored. Stalnaker also stressed that the grade sheets must be removed from the answer books and that new ones should be attached. He stated that "this procedure gives a _second and completely independent reading_ of the paper and provides a means of checking the reliability with which the papers are read" (p. 672). Although these are common practices utilized in scoring large-scale English examinations today, the importance of these techniques for maintaining student anonymity and helping to reduce marker bias had not been specifically documented before Stalnaker's study.

Cast's research (1939; 1940) was the first to compare the reliability of different ways of grading writing. His study had markers use four distinct methods of scoring--their individual method, the achievement-of-aim method, the general impression method, and the analytic method. After forty compositions written by fourteen- and fifteen-year-old girls had been graded, Cast concluded that analytic scoring was the most reliable technique to use when grading written compositions. He summarized his research findings by stating:

> Of the four methods of marking employed, the "analytic" method (allotting separate marks for specified points or qualities), though laborious and unpopular, appears almost uniformly the best, as judged by each of the criteria. The method of marking by "general impression" discriminates more widely among the individual candidates, but tends to judge them by more superficial characteristics (p. 59).

A study conducted by Anderson and Traxler (1940) attempted to confirm the results of Stalnaker's earlier research (1934). Based upon Stalnaker's findings of improving scorer reliability through the training of markers, Anderson and Traxler conducted a reliability study in June, 1934. They had more than one hundred high school students write essays that were scored by two graduate students in English from the University of Chicago. The correlation between the scores of the two readers was 0.894. The authors decided "to undertake another study of a wider scope in order to verify the findings obtained earlier. . ." (1940, p. 524). Two essays were written by 281 high school students. The first essay was completed at the beginning of a school year and the second almost a year later. The compositions were not scored immediately, but rather were filed for future study. Later in the year, a graduate student marked the essays by following a set of detailed instructions which had been provided. The researchers inserted a number of compositions which had already been scored into the pile of essays which were still to be graded. From the 138 papers that were rescored by the graduate student, a correlation of 0.893 was obtained. The writers concluded that "a single reader can be trained to read English essay tests with high reliability" (p. 526). When two readers graded the same set of papers, a correlation of 0.878 was obtained. Anderson and Traxler reinforced their findings by stating:

> The data in the two studies and those reported earlier by Stalnaker provide strong evidence that it is possible to train a group of readers who will not vary greatly among themselves in the scores which they assign to essay tests in English (p. 526).

By the beginning of the nineteen fifties, the problems linked to reliability had been narrowed down to difficulties associated with several markers scoring a paper, a single marker re-scoring an examination, and the problem of a student's written work varying according to assignment and daily performance. Research in this decade concentrated on improving reliability when these factors were considered.

Finlayson (1951) conducted a study which was concerned with both intra- and inter-rater reliability. Two months after his raters had scored essays written by twelve-year-old students, they were sent the same essays for re-scoring. A reliability of 0.786 was reported for the study and Finlayson concluded:

> It will be seen at once that, while the performance of each marker is similar on both essays, there is a considerable difference in level and scatter of the marks from marker to marker (p. 128).

Coward's study conducted in 1952 dealt with a slightly different aspect of reliability in that she compared the atomistic and holistic methods of scoring. Based upon one hundred themes that were scored by sixteen markers, Coward concluded:

> The wholistic [sic] method of reading employed was considerably faster than the atomistic method. It is likely that the reading reliability of the two methods would be about the same if the same amount of time were taken (p. 93).

The researcher commented upon reader variation when either atomistic or holistic scoring was used and she expressed concern regarding her findings by stating:

> There may be an intrinsic difference in what is measured by the atomistic and wholistic [sic] methods of grading; but, if there is, it was not sufficiently great to overshadow the individual variations between readers which occurred regardless of the method of reading (p. 91).

Research conducted by Huddleston (1954) and by Vernon and Millican (1954) found a lack of consistency in inter-rater reliability. Huddleston obtained coefficients of 0.68 and 0.62 in two separate studies and made the comment:

> The results are discouraging to those who would like to develop reliable and valid essay examinations in English composition--a hope that is now more than half a century old (p. 204).

Vernon and Millican obtained correlations of 0.415 when fifteen markers graded more than two hundred essays written by students attending a teachers college. They were able to explain the low correlation by breaking down their findings into several key components and concluded:

> Inadequate correlations between different markers of the same essays chiefly occur when the candidates are selected and therefore homogeneous in ability. They are also lowered when the writers are mature, the essays are short, or the markers relatively inexperienced. A still more serious

> source of inconsistency in assessing English ability is the varying performance of candidates when writing essays on different topics (p. 73).

Vernon and Millican also made an interesting statement about markers and less able students. They found that "the less skilled or mature the writers, the more their essays are marked for mechanics rather than for the relatively subjective qualities of style and thought-content" (p. 66).

Wiseman (1956) found that reliability increased when a team of markers was used in grading papers and he stressed that unreliable readers should be eliminated from a scoring project. He expressed concern that very few inconsistent graders were dropped from marking sessions in England and he pointed out that reliability could be increased if inconsistent graders were dismissed.

The topic of reliability of scoring was stressed during the nineteen sixties as English teachers continued to struggle with the problem of consistency in grading. Perhaps the best known reliability studies were conducted during this decade by Diederich, French, and Carlton (1961) and by Braddock, Lloyd-Jones, and Schoer (1963).

Diederich, French, and Carlton (1961) examined the scores that were awarded to three hundred compositions written by college freshmen. The markers consisted of English teachers, social scientists, natural scientists, writers and editors, lawyers, and business executives. In a discussion of the study one of the researchers stated:

> . . .it is probably typical of the amount of disagreement one would find in any large group of

> readers without such training and discipline that, out of the 300 essays graded, 101 received every grade from 1 to 9; 94 percent received either seven, eight, or nine different grades; and no essay received less than five different grades from these fifty-three readers (Diederich, 1974, p. 6).

The correlation obtained from the scoring data was 0.31; however, it was determined that, even with this low correlation, certain characteristics caused the markers to agree with regard to what was important in the grading of the essays. The authors found that readers were most influenced by the ideas expressed in the compositions and that they were next most influenced by the errors the students made in "usage, sentence structure, punctuation, and spelling" (Diederich, 1974, p. 7). The third most important grading criterion consisted of "organization and analysis" (1974, p. 8), the fourth was "wording and phrasing" (1974, p. 8), and the last was composed of the "personal qualities revealed by the writing" (1974, p. 8). The writers labelled this component "flavor" (1974, p. 8). These five factors made up the components of the scoring models in later research conducted by Diederich.

The report that Braddock, Lloyd-Jones, and Schoer produced for The National Council of Teachers of English (1963) contained many references to reliability studies completed over the years. The writers synthesized the findings of others in a discussion of composition rating. They mentioned that an evaluator could never be sure that "the student is fully using his ability" (p. 6) when writing and they warned that the writer variable was important when assessing written performance. They stressed the importance of the assignment variable and emphasized that "the topic, the mode of discourse, the time afforded for writing, and the examination situation" (p. 7) must be controlled. The writers recognized that the rater variable was important in that a marker would "vary in his own standards of evaluation" (p. 10). To reduce these difficulties two cautions were given. The first dealt with rater fatigue, and the researchers stated:

> Fatigue may lead raters to become severe, lenient, or erratic in their evaluations, or to emphasize grammatical and mechanical features but overlook the subtler aspects of reasoning and organization. Consequently, raters should not be permitted to rate late at night or for lengthy periods during the day, and they should have regular rest periods to help them maintain their efficiency (p. 11).

The second concern dealt with proximity, and the writers found:

> . . .it seems highly desirable to have all of the raters working in the same or adjoining offices, where the investigator can be present and, without entering into the rating himself, insure that everything runs smoothly (p. 11).

Another component with which the authors were concerned was labelled the colleague variable. It was recognized that markers varied in the way they evaluated writing and that consistency was an important factor in order to maintain some degree of fairness towards the writers of the compositions. The authors recommended:

> A common set of criteria seems essential in coping with the colleague variable; if raters are not evaluating for the same qualities, they cannot be expected to rate with validity or reliability. Three principal means of achieving this commonality

> are composition scales, a "general impression" method of rating, and an "analytic method" (p. 12).

It was also suggested that writers should be given the opportunity to practice rating papers before the actual scoring began and that these ratings could be used "as a basis for pairing raters with differing standards of severity and leniency" (p. 15). A warning was made against stressing mechanics over the other factors involved in scoring compositions and that raters should be careful not to emphasize one factor and ignore the others.

In a study by Britton, Martin, and Rosen (1966), the "rater and colleague variables" discussed by Braddock et al were again recognized. The authors believed that it was necessary to "take a constructive interest in the differences between examiners" (p. 31) and that "if the judgement upon a piece of writing is to do it justice the marker must respond to it in a more personal way, a way that reflects his outlook, his attitudes, his personality" (p. 10). The authors tentatively proposed:

> . . .examiners, where they differ, differ in the areas of their most sensitive discrimination and . . .this is the very element in their judgement that we should wish to incorporate into our assessment (pp. 10-11).

Britton et al recognized that the concern about individual differences in the scoring of compositions needed further research.

In another study conducted in 1966, Myers, McConville, and Coffman discussed the reliability of scoring as it was calculated in a large-scale study on a daily basis. They found that inter-rater reliability dropped markedly (from 0.732 to 0.589) on the final day of scoring in a large-scale essay examination situation in which more than eighty thousand compositions were read over a five-day period. The writers concluded their research with the following questions and remarks:

> Is it possible to maintain a high state of vigilance at the end of the reading period? If the readers are mature and conscientious people, as they were in the present setting, they might resolve the problem themselves if it were pointed out to them that there was a tendency for them to slip at the end of the reading period. If, however, the readers were not able to maintain a state of vigilance by their own efforts, then the reliability problem will appear more troublesome.
>
> We have been assuming that the drop in reliability occurs at the end of a reading period because the readers are anticipating the completion of their task. If this is so, it means that there would be an equivalent drop in reliability regardless of how long the reading period was. This implies immediately, of course, that this problem cannot be handled by simply shortening the reading period by any small amount. It would seem that some external source would be needed to bolster the reader morale and effort (p. 53).

In a study published by Godshalk, Swineford, and Coffman (1966), specific scoring procedures designed to improve the inter-rater reliability were presented in great detail. Their markers were asked to rate compositions holistically, awarding grades from one to three. A "two" paper was considered to be an average piece of writing while a "one" paper was inferior and a "three" was superior. In awarding grades, the markers were asked "to judge each paper on its merits without regard to other papers on the same topic" (p. 10). They were told to use the full range of marks and that awarding all or almost all average marks (the grade of two) should be avoided. The authors attempted to improve the reliability by adapting the following procedures:

> Standards for the ratings were established in two ways: by furnishing each reader with copies of the sample essays for inspection and discussion, and by explaining the conditions of administration and the nature of the testing population; and by having all readers score reproduced sets of carefully selected sample answers to all five questions and to report the results. The scores were then tabulated and announced. No effort was made to identify any reader whose standards were out of line, because that fact would be known to him and would be assumed to have a corrective effect. The procedure was repeated several times during the first two days of scoring to assist readers in maintaining standards (p. 10).

After following these techniques, the researchers found that the reliability of the scoring increased; however, they had a number of other important recommendations to make in order to raise the reliability even further. They found that reliability decreased when students were given the freedom to write on a choice of topics and they thus suggested that "no options should be permitted" with regard to the assigned topic. They found that reliability increased when a four-point rather than a three-point scale was used, and they favored the utilization of the four-point marking scale for holistic scoring. They summarized their findings by stating, "the reliability of essay scores is primarily a function of the number of different essays and the number of different readings included" (p. 39) and concluded that, "the most efficient predictor of a reliable direct measure of writing ability is one which includes essay questions or interlinear exercises in combination with objective questions" (p. 41).

Follman and Anderson's study (1967) reported a high inter-rater reliability, from 0.810 to 0.953, when five different methods were used to score ten themes. All of the twenty-five markers were students enrolled in an English methods course and the researchers commented that the similar training background of the students may have caused the high reliability. They stated:

> When a group of heterogeneous raters uses an evaluation system and the mean reliability is higher than what it would be without the system, it appears that the system provides a sensitizing to certain elements of a theme and to certain values used in theme evaluation (p. 199).

The writers concluded that the improvement of reliability depended upon sensitizing heterogeneous raters so their judgments were guided more carefully.

Similar findings were cited in a study by Smith (1969). The researcher mentioned Follman and Anderson's idea of the similarity of the background of the raters as a contributing factor in increasing reliability and he agreed that "teachers in one large school district and undergraduate and graduate students in a few classes in one university could certainly be considered more homogeneous than a general population" (p. 192).

The nineteen seventies brought many more studies dealing with the topic of essay scoring and reliability. These studies shied away from the earlier works in that the research was less concerned with reliability coefficients than it was with making general recommendations for markers.

In a study conducted by McColly (1970), the issue of competence was brought forth. He believed that competence was the most crucial marker characteristic and elaborated on the term by stating, "the kind of competence meant here is scholarship or knowledgeability. Some normalcy of temperament is required as well for the demonstration of competence in essay-rating. . ." (p. 150). McColly discussed the importance of training and orientation for markers and stated that two methods could be used to train readers effectively. The first presented markers with "predetermined standards and criteria and some kind of artifact for their application, such as exemplary essays" (p. 150), and the second called the readers together to discuss the reading, examine a series of compositions, and arrive "at a determination of their own standards and criteria through consensus" (p. 150). McColly also stressed one of Stalnaker's earlier findings (1934), that of the importance of practice in raising the level of reader performance. A final point made by McColly dealt with the matter of marker fatigue, a concern that had also been examined by Coffman, McConville, and Myers (1966). McColly recognized that fatigue was a problem in scoring; however, he pointed out that researchers could not be certain when it began to be felt and what the consequences might be.

Coffman (1971b) discussed the essay examination and listed four ways of improving reader performance. He found that the teacher could reduce rating error as follows:

> First, use a sufficiently fine scale for recording the ratings. Second, develop clear reference points to anchor the scale. Third, distribute the error randomly rather than systematically. Finally, include multiple rating where feasible (p. 33).

Akeju's research (1972) differed from most other reliability studies in that it involved high school compositions written in Ghana. Akeju discussed the problems of inter-rater and intra-rater reliability in essay scoring and agreed with Gulliksen (1950, p. 212) and Diederich (1974, p. 33) in stating that an acceptable reliability coefficient was 0.80 or better. He concluded that the reliability coefficient of 0.72 that was reported from his study of one hundred essays was unsatisfactory and that to raise the reliability it would be necessary to increase "the number of markers who read each essay" (p. 179) and to adjust statistically the "differences in standard between examiners" (p. 179). This systematic adjustment of standard differences had not previously been discussed as a solution to the problem.

The publication Measuring Growth in English by Diederich (1974) made a number of suggestions for improving reliability. Early in the book, the author recognized the difficulties associated with essay scoring and suggested that one way to improve the reliability of an examination in English was to include written compositions as well as "a few sections of objective items on related parts of proficiency in English" (p. 2). Diederich preferred two essays, one written in the morning and another written in the afternoon to allow for the writer variable and he "encouraged the teachers to work rapidly and to trust their first impressions, since we found that this increased the reliability of grading" (p. 3). The researcher supported the scoring of essays by two markers with a third rater examining the marks when the scores differed by more than one grade point. He then suggested:

> Papers on which the two grades differed by more than one full grade-point were referred to a small committee of the most experienced and trusted readers, who did not know what grades these papers had received; they knew only that the grades differed. One member of this committee would give each paper a third independent reading, and a clerk would substitute this grade for whichever of the two previous grades was farther from it. If they were equally distant, he discarded the grade nearest the mean, since combining or averaging grades pushes everybody toward the middle, and we want to keep them spread out as far as possible. But if the first two grades were B and D and the third was C, he discarded the lowest grade to give the student the benefit of the doubt (p. 20).

Diederich's text consolidated earlier reliability research effectively and demonstrated to teachers that simple techniques could be used to increase reliability within large-scale testing situations and within their own school's marking program.

An article by Cooper (1975) reiterated the importance of using student identification slips to preserve anonymity, the necessity of having teachers practice using whatever scales had been devised, and the significance of a committee to examine papers with more than a ten-point difference between raters. Cooper added one new factor in grading essays by a number of markers, and that was the concept of the rater number. The researcher stressed the importance of having each teacher assign "himself any random three-digit number as his Reader Number" (p. 119). This was done to keep the rating process anonymous.

Horner's article (1978), like Cooper's, emphasized the importance of practice scoring and of writer and reader anonymity when compositions were marked by a number of raters. She suggested that discrepant scores could be "noted in the final scoring procedure, during which time the scores are compared and a majority decision reached" (p. 60) or that an alternate procedure would involve "the use of a 'master reader' who makes decisions on controversial papers" (p. 60). Horner recognized fatigue as a factor but was more specific than the other researchers in that she recommended ten minute breaks for every hour of scoring.

Cooper and Odell (1977) made a number of suggestions based upon the findings from the earlier research of Stalnaker (1934), Follman and Anderson (1967), Coffman (1971), and upon their own ideas about composition scoring. They summed up the research of the decade when they stated:

> The raters themselves must be carefully trained. They should practice using the holistic scoring guide with sample papers exactly like those they will be scoring, and they must be able to make their judgments within the context of the range of performances in the particular set of papers they are scoring. Scoring is always relative to the set of papers at hand and must take into account of [sic] the writing task, the conditions under which the writing was done, the age and ability of the students, and the full range of quality of the papers. Reliability cannot be achieved when some raters are using an absolute standard of quality, perhaps that of published adult writing. Some

> papers must receive the highest scores, some the lowest, and most the scores in the middle range. Raters should check the reliability of their ratings during training to insure they reach an acceptable level before they begin the actual scoring. Then once the scoring is underway, they should periodically check themselves on perhaps every twenty-fifth paper. . . (p. 20).

The nineteen eighties continued the trend of the decade before it by producing few actual reliability studies based upon writing samples. Instead, the research most commonly made recommendations for improving scoring reliability applicable to teachers, school districts, and larger divisions when essay testing was carried out.

Miles Myers (1980) published a book which was devoted to the topic of writing assessment and the grading of compositions. He supported the use of holistic scoring when marking compositions and discussed the importance of scoring criteria and anchor papers. Myers raised the issue of "slippage" within a scoring category and stated that "although several papers might have one number, they often will have recognizable differences" (p. 31). He also labelled a group of papers which did not fit neatly into one numerical value as "splitters" and gave his remedy for the difficulty when he emphasized:

> The solution for splitters, then, is to clearly define the difference between upper-and lower-half papers and to define what should be emphasized on either side of the splitter. When a question exists about a particular paper, examine the anchors on both sides of the suggested score (p. 38).

The author stressed the importance of helping deviant scorers by table leaders through a procedure whereby the leaders conferred with deviant markers and showed how the paper could be fitted within certain boundaries. It was also suggested that if "an individual reader continues to give odd scores, the table leaders should assign that reader to the head reader for special duties, such as collecting papers and counting them" (pp. 43-44). This was the most radical method cited for improving reader reliability in the review of the literature.

An article by Odell and Cooper (1980) introduced the term "range finders." The authors elaborated upon the range finder procedure as follows:

> First a scoring coordinator and several assistants read through the complete set of papers, identifying "range finders," papers which illustrate the full range of performance on the assignment. Assuming that this range might be represented along a four point continuum, the leaders of the evaluation identify pairs of papers which, in their judgment, illustrate the distinction between a 1 paper and a 4 paper, between a 2 paper and a 3, and so on. During their own reading, the coordinator and his or her assistants discuss the reasoning underlying their judgments about the quality of papers. But in training scorers to evaluate the papers, these leaders do not prepare a list of criteria for distinguishing between better and poorer papers. Rather, they choose papers that illustrate these criteria. The scorers' job is to infer these criteria, with the leaders providing only brief explanations as to why they rank the papers as they do (p. 36).

By following this technique, graders were given guidelines regarding the marking standards, yet they were still able to make the final decisions regarding the marks themselves. Braddock et al (1963), Godshalk et al (1966), Follman and Anderson (1967), and Smith (1969) also believed that scorers should feel that they had made important decisions relative to the marking process.

Lees (1981) listed seven ways to improve reliability in essay evaluation. These included:

1. Raters with a common background are chosen.
2. Raters pass a qualifying test in order to be eligible to read.
3. Raters agree to share a set of common evaluative criteria, either criteria they have formulated themselves or criteria an investigator supplies.
4. Raters study a common set of scored sample essays before beginning to evaluate on their own.
5. Raters practice reading unscored sample essays together.
6. Raters discuss differences that arise in their evaluations of sample essays.
7. Raters are monitored during the actual evaluation session to ensure that they're reading consistently (pp. 3-4).

Nyberg and Nyberg (1982b) demonstrated that essay models could be used to improve scorer reliability in marking writing. They found that compositions "should be graded so as to produce separate scores for mechanics and style-content" (p. 14) and that the holistic approach "was more efficient than the 'atomistic' approach" (p. 14). This latter conclusion was based upon a correlation of 0.97 found between the "true" scores for the variable "general impression" and the "true" total scores for the essays. The "true" scores were means based upon twenty-one independent readings of the papers by different markers.

Meredith and Williams (1984) addressed the concern of writing assessment in large-scale testing programs and were able to summarize many of the factors important to increasing scoring reliability. The authors first mentioned the importance of ensuring that "the scoring criteria are understood and applied appropriately by all raters throughout the scoring sessions" (p. 12). Ensuring understanding of criteria could most easily be done, they found, by developing a scoring guide which "contained materials that are relevant to accurately applying the scoring criteria" (p. 12). Guidelines were given for the training of scorers and it was suggested that this be carried out through the scoring of sample papers which had already been premarked. Raters were to be given the "opportunity to become familiar with the scoring guide so that [they] can acquire a feel for the scoring criteria" (p. 13) and once familiar, they were to score "solid" papers, representing definite mark values agreed upon by the prescoring committee and "marginal" papers, representing problems in grading. It was suggested that markers should be monitored on a daily basis by practicing on papers that had previously been scored by a range-finding committee. They also suggested the following:

> A complete quality control system would use daily reports on accuracy and consistency, as well as to-date indices of rater performance. In addition to accuracy data, information contained on this monitoring report would include indices such as the total number of readings, the number and percent of pass/fail papers; the number and percent of scorer agreement with other scorers; correlation coefficients; percent and direction of score point discrepancies; number and percent of papers needing resolution; and number and percent of agreements with resolution scores (p. 14).

Gee (1987) conducted a study based upon compositions written at the high school leaving level in 1978 for the Alberta English Written Composition Achievement Test. While not a reliability study, the research had implications relevant to this area. Gee examined the topics that had been assigned and concluded:

> . . .topics cannot be selected, despite the care and rigor exercised in their selection, so that they are of equal difficulty or quality. The structuring of the topic statements themselves, and the differences in that structuring, appear to influence the development of an essay. It also appears that, where several topics are made available on an essay examination, the students who opt for the least popular topics tend to make the highest grades or scores. Markers may be biased against a topic, get bored when marking the same topic over and over again, or more capable students may opt for one to two particular topics. Even the order of topics may influence their equality (p. 112).

Summary of Literature and Extent of Application By Alberta Education

Most of the reliability studies which dealt with the subject area of English language arts contained common elements. These studies conducted from the early nineteen hundreds to the present date have been summarized in Table II-1.

When Alberta Education initiated its diploma examination testing program for English 33 in January, 1984, many of the recommendations from the reliability studies conducted in the past ninety years were put into practice. The Student Evaluation Branch chose to follow some of the suggestions made by the various studies rather closely while recommendations made by other researchers were virtually ignored. Policies which were adopted are listed in Table II-2; those which were not followed are given in Table II-3.

Table II-1

SUMMARY OF RESULTS OF RELIABILITY STUDIES

Date & Researcher(s)	Rel.	No. Rdrs.	Level	Comments
1912 Starch & Elliott	varied	326	H.S.	Graded by teachers English students, and measurement students
1922 Darsie	0.88	18	Gd. 4-8	Used essay models; marks on papers
1934 Stalnaker	0.88	pairs	Univ. Entr.	Read twice by same reader
1937 Stalnaker	0.84	100	CEEB	Papers read by two readers
1939, 1940 Cast	0.49	12	Age 14-15	Analytic method judged to be best
1940 Anderson & Traxler	0.89	2	Grade 11	Two forms of test given
1940 Anderson & Traxler	0.88	2	H.S.	Markers followed written directions
1951 Finlayson	0.79	6	Age 12	Re-scored after 2 months; general impression method
1952 Coward	0.70	16	Adult	Holistic and atomistic procedures compared
1954 Huddleston	0.68 0.62	? 39	H.S. H.S.	First study used to set up next one
1954 Vernon & Millican	0.42	15	Educ.	Papers read twice by two markers
1956 Wiseman	0.80	4	Primary	Two papers written four months apart
1961 Diederich, French	0.31	53	Univ.	Readers from five fields; no guidelines were imposed
1966 Britton, Martin & Rosen	0.51	100	O Level Exams	General impression method; scored at least three times

Table II-1 (Cont'd)

Date & Researcher(s)	Rel.	No. Rdrs.	Level	Comments
1966 Myers, McConville & Coffman	0.73	145	CEEB	Holistic scoring; reliability drop on final day
1966 Godshalk, Swineford & Coffman	0.71	25	Grade 11-12	Holistic scoring; each paper read by five markers
1967 Follman & Anderson	0.81-0.95	25	Grade 5	Five methods of evaluation used
1969 Smith	0.60-0.79	218	Grade 5	Two compositions scored four weeks apart; holistic
1970 McColly	0.43	16	H.S.	Four point scale; fast marking (one essay per minute)
1972 Akeju	0.72	7	H.S. Leave	Low reliability because of "inherent idiosyncracies of scorers"
1982 Nyberg & Nyberg	0.78 0.77	6	H.S. Leave	Mechanics & style-content rated by use of models
1984 Marsh & Ireland	0.70	6	Grade 7	Two ratings made ten months apart; holistic, analytic
1985 Swartz & Whitney	0.85	3	H.S.	Holistic method; 2 reader minimum and third resolved problems
1985 Sweedler-Brown		26	Univ.	Analytic and holistic; two readers and third resolved differences

Table II-2

RECOMMENDED PROCEDURES ADOPTED BY ALBERTA EDUCATION

Recommended Procedure	Researcher(s) and Year
Practice scoring on sample papers and train markers.	Stalnaker (1934), Anderson & Traxler (1940), Godshalk et al (1966), McColly (1970), Cooper (1975), Cooper & Odell (1978), Horner (1978), Lees (1981), Meredith & Williams (1984).
Instruct scorers in small groups.	Stalnaker (1934).
Conceal students' names and place no marks on essay papers.	Stalnaker (1937), Cooper (1975), Horner (1978).
Assign a number to each marker.	Cooper (1975), Horner (1978).
Use a combination of holistic and analytic methods of scoring.	Cast (1940), Coward (1952), Godshalk et al (1966), Braddock et al (1963), Diederich (1974), Myers (1980).
Compile and issue a daily report card to the markers.	Meredith & Williams (1984).
Score each paper more than once.	Coffman (1971b), Akeju (1972), Diederich (1974).
Train the markers, and have them working in close proximity to one another.	Braddock (1963).
Utilize a common set of criteria for scoring the papers.	Braddock (1963), Follman & Anderson (1967), McColly (1970), Coffman (1971b), Lees (1981), Meredith & Williams (1984).
Do not identify deviant markers to the others.	Godshalk et al (1966).
Limit the choice of topics available to the students writing the test.	Godshalk et al (1966).
Allow for fatigue.	Braddock et al (1963), McColly (1970).

Table II-2 (Cont'd)

Recommended Procedure	Researcher(s) and Year
Appoint homogeneous leaders; that is, leaders who have taught the course.	Smith (1969), Lees (1981).
Require students to write more than one essay.	Diederich (1974).
When grades differ, assign a marker the task of making the final decision.	Diederich (1974), Cooper (1975), Horner (1978).
Recognize and give special attention to "splitter" papers, that is, papers that are assigned scores near the cutting points.	Myers (1980).
Utilize "range finding" procedure, that is, identify a set of papers that represents the full range of performance.	Odell & Cooper (1980).
Train markers on papers representing a wide range in quality of writing.	Meredith & Williams (1984).

Table II-3
RECOMMENDED PROCEDURES NOT ADOPTED BY ALBERTA EDUCATION

Recommended Procedure	Researcher(s) and Year
Assign deviant markers to other tasks.	Wiseman (1956), Myers (1980).
Employ scale with 7 or more points.	Coffman (1971b).
Periodically, check work done by markers after work commences.	Cooper & Odell (1977), Lees (1981).
Distribute marker error randomly by having different readers score each question.	Coffman (1971b).
Adjust markers scores so as to correct for differences in grading standards.	Britton et al (1966), Akeju (1972).
Have students write two essays on different occasions.	Diederich (1974).
Require the raters to pass a qualifying test.	Lees (1981).

CHAPTER III

DESIGN OF THE STUDY

General

The overall purpose of the study was to address the problem of essays that resulted in discrepant scores, and to calculate the reliability of an essay examination as a whole. The essays that posed difficulties in scoring were studied with respect to four sources of variations in marking--those related to the compositions, those related to the readers, those related to the scoring variables, and those related to the writing tasks. Reliability of the essay test was calculated, as opposed to the usual task of computing the reliability of scoring. The procedure was to collect samples of compositions, to gather data on scores awarded to those compositions, and to conduct an analysis focusing on the specific objectives of the problem.

In order to secure relevant data, the cooperation of a large-scale scoring authority was necessary. This cooperation was readily provided by the Student Evaluation and Records Branch of Alberta Education and compositions written under examination conditions in January, and June, 1986, (part A of the examinations) were made available.

Sampling

Three samples were drawn. The first consisted of examinations which were selected for use in the scoring consistency checks (or reliability reviews as they were called). This sample was not intended to be representative of any population of papers submitted. These papers were chosen by having a group of trained readers pick papers which they believed were useful in demonstrating the various situations that markers might encounter when they scored the examinations. Each of these papers was discussed by the trained readers and a consensus was reached before the paper was finally placed in the reliability review sample. The sample consisted of seven compositions submitted in January, 1986. The second sample, drawn by a similar procedure from examinations written in June, 1986, was composed of a total of ten compositions. The two reliability review samples were not drawn merely for the purposes of this study. Rather, they were intended for use in training the groups of readers for consistency. During the marking session, the papers were photocopied and distributed one at a time to all of the markers. These papers were utilized during the reliability review checks which took place every morning and every afternoon.

The third group of papers consisted of a random sample of ninety-six test papers which had been taken from the January, 1986, scoring session. The actual procedure involved the selection of examinations on the basis of computer-generated random numbers which had been applied to the identification numbers printed on the test booklets. Unlike the two reliability review samples, the third sample was drawn for use in this study, only. One photocopy of each paper was made for the purposes of this study.

Data Gathered

Each of the reliability review papers was scored by all of the readers engaged in marking at each of the two sessions. That is, for the seven papers in sample one, each had been graded by all seventy-five readers on every variable, and for the ten papers in sample two, each had been similarly graded by all sixty-four markers on every variable. The result, then, was seventeen matrices of scores as represented in Table III-1.

The random sample of ninety-six papers (the reliability sample) was scored in the same manner that was employed for the total group of 4218 papers (Student Evaluation, 1986, p. 9). This procedure involved having each paper graded independently by three of the seventy-five markers. The marking was based upon the same set of scoring descriptors listed in Appendix A. No systematic procedure was used for deciding which set of three markers would score a particular paper; instead, the selection of groups of papers by the markers themselves was left to chance. The total group of examination papers was divided into several large batches. Each batch was subdivided into bundles of eight papers, by clerks, then was submitted for scoring. Only one batch at a time was scored. The bundles of eight were circulated until each had been graded three times. After completing the scoring of a paper, a reader recorded his or her marker number in a box printed on the back of each paper. A reader,

Table III-1

COLLECTION OF RELIABILITY REVIEW DATA

January, 1986, Reliability Review Session

Paper #1

Task	Essay Variables	Reader #						
		1	2	3	-	-	-	75
I	Thought & Detail	X(1,1)	X(1,2)	X(1,3)	-	-		X(1,75)
I	Organization	X(2,1)	X(2,2)	X(2,3)	-	-		X(2,75)
I	Choice							
I	Convention							
II	Thought & Detail							
II	Organization							
II	Writing Skills							
III(1)	Thought & Detail							
III(1)	Writing Skills							
III(2)	Thought & Detail							
III(2)	Writing Skills	X(11,1)	X(11,2)	X(11,3)	-	-		X(11,75)

Note: X(1,1) was the score awarded in row 1, column 1, that is, the score given for Thought and Detail by Reader #1.

This format continued through paper #7. A similar set of ten matrices resulted from the July, 1986, scoring session.

when choosing a bundle, would ensure that he or she had not graded that set at an earlier time during the scoring session by checking the numbers in the boxes.

Reliability Review Procedures

A reliability review check consisted of dividing the markers into groups of approximately five, including a leader, and then distributing copies of one sample paper to all readers. The paper was then scored in the usual manner after which the marks for each variable were tallied and a summary table was constructed. Readers whose scores were discrepant were expected to explain their grades, although observations and ideas were welcomed from any member of the group. At this point the markers were invited to change their grades, but there was no direct pressure on them to do so. Each group leader then entered marks on tally sheets based upon the final consensus. Next, summaries of the scores recorded by the groups were drawn up and posted; that is, a table showing the distribution of grades for each variable on the initial reading was constructed, as was a table showing the final totals after discussion and after changes had been made by individual markers.

This procedure ordinarily took an hour to complete, although the time lost was much less because readers returned to the task of scoring papers as soon as their group had completed the reliability review session. These scorers continued marking papers while tables were being constructed by the clerical staff members. The posting of the reliability review tables was announced a short time after the reliability review session, and each reader then took the opportunity to compare his or her grades with the final consensus. Group leaders also compared their table's scores with those on the overall table.

The primary focus was upon how each individual performed in relation to the total group. No direct suggestion was made that a reader was grading too easily, too severely, or inconsistently; however, it was generally assumed that the markers would see this for themselves and adjust their standards. Experience had shown that some of the scorers would be inflexible.

During the first day of the scoring session, a reliability review session took place once during the morning and once during the afternoon. In general, the morning session involved comparing grades with markers from the group in which a person normally worked; in the afternoon session, group leaders remained at their assigned tables and the members of the Student Evaluation and Records Branch Staff assigned markers to new tables on a random basis. By following this procedure, markers were given the opportunity to work with many individuals and to compare their performance with that of others. Toward the end of the scoring session, reliability reviews were generally conducted once a day rather than twice daily because it was assumed that markers were becoming more familiar with the different types of papers as they gained greater experience.

For the purposes of this study, data from the reliability review samples were taken from the initial-marking rather than the post-marking discussion. This procedure was used because it was felt that the initial scores more truly represented the actual scores being awarded by the readers.

Details of the Design

Data from the two review samples have been represented in five tables to assist the reader in following the descriptions of the various analyses. Table III-1 is a master table representing all data collected relative to the review samples. The data were entered on punch cards so that each contained one reader's scores for a single essay. Each column in Table III-1, therefore, represents one card.

Table III-2 shows how each of the raw scores, awarded by a single reader, was weighted to produce the scores actually used in computing students' marks. This was carried out according to the guidelines established by Alberta Education when the examination was brought into effect in 1984.

Table III-3 represents a matrix of total scores, weighted as outlined in Table III-2. Values in each row consist of total scores awarded for a single essay. The ideal situation would be ~~for~~ the scores in each row to be identical, and therefore to have standard deviations of zero. Each column consists of scores awarded by a single reader. Ideally, the means should be equal, because each reader had scored the same set of papers. Differences in means

Table III-2

English 33 Diploma Examination Blueprint

Description of the Writing Assignment	Reporting Category	Proportion of Total Mark
Personal Response to Literature		
The student is required to read a short literary selection that serves as a stimulus for a personal or reflective response.	1. Thought & Detail	10 %
	2. Organization	5 %
	3. Matters of Choice	5 %
	4. Matters of Convention	5 %
Functional Writing		
The student is required to write for a specified purpose and audience.	1. Thought & Detail	10 %
	2. Organization	2.5%
	3. Writing Skills	2.5%
Response to Visual Communication		
Writing assignments are connected to a cartoon. The student is required to write about main ideas, techniques of communication, and personal reactions.	1. Thought & Detail	5 %
	2. Writing Skills	5 %
		50 %

(Student Evaluation, 1986a, p. 10)

Table III-3

WEIGHTED SCORES BY READER AND PAPER

(January, 1986, Review Sample)

			Reader No.			
Paper No.	1	2	3	--------	75	S.D. for Papers
1	X(1,1)	X(1,2)	X(1,3)	---	X(1,75)	SP(1)
2	X(2,1)	X(2,2)	X(2,3)	---	X(2,75)	SP(2)
3	X(3,1)	X(3,2)	X(3,3)	---	X(3,75)	SP(3)
-	-	-	-		-	-
-	-	-	-		-	-
7	X(7,1)	X(7,2)	X(7,3)	---	X(7,75)	SP(75)
Means for Readers	$\bar{X}(1)$	$\bar{X}(2)$	$\bar{X}(3)$		$\bar{X}(75)$	
S.D. for Readers	SR(1)	SR(2)	SR(3)	---	SR(75)	

Note: Entries in the table were weighted total scores for papers. Raw scores awarded by readers were weighted as indicated in Table III-2. For example, X(2,1) was the weighted total score awarded to paper 2 by reader 1.

reflected differences in scoring standards by the readers. The standard deviations, in an ideal situation, should also be equal; variations reflected differences in how readers tended to spread the scores.

Table III-4 is a matrix representing correlations of total weighted scores for essays for all pairs of readers. Mean correlations for each reader are represented at the foot of each column.

Table III-5 is similar to Table III-4 in that correlations between all pairs of readers are represented. Table III-5, however, deals with only one scoring variable at a time. The complete table, then, consists of eleven sections, of which only the first is illustrated. The mean correlations, shown at the base of the columns, are indicators of the degree of agreement among readers.

I. Compositions Having Low Scorer Reliability

A. Problems Related to the Compositions

Problem papers were identified as those having the largest discrepancies of scoring amongst the readers (see Table III-3). Standard deviations, calculated for all essays across all readers, were used as measures of scoring discrepancy. That is, each essay was scored by more than sixty readers, and the standard deviation of the scores gave an indication of the disagreement. It should be noted that in the actual scoring situation, the raw scores were weighted before being combined (see Table III-2). These weighted scores were summed, and then used in calculating the standard

Table III-4

CORRELATIONS BETWEEN TOTAL SCORES FOR ALL PAIRS OF READERS

(January, 1986, Reliability Review Sample)

Reader No.	Reader No. 1	2	3	--------	75
1		r(1,2)	r(1,3)	---	r(1,75)
2	r(2,1)		r(2,3)	---	r(2,75)
3	r(3,1)	r(3,2)		---	r(3,75)
-	-	-	-	-	-
-	-	-	-	-	-
75	r(75,1)	r(75,2)	r(75,3)	-	
Mean Correlation	$\bar{r}(1)$	$\bar{r}(2)$	$\bar{r}(3)$	-	$\bar{r}(75)$

Note: Correlations between any two readers could be identified. For example, r(2,1) was the correlation between total scores over the 7 essays awarded by readers 1 and 2. All correlations appeared in the table twice, because the value labelled r(1,2) was the same as that labelled r(2,1), etc.

No entries were made in the spaces where the two reader numbers would be the same, for example r(1,1). This would be the correlation between reader 1 and reader 1.

Table III-5

CORRELATIONS BETWEEN READERS FOR ALL SCORING VARIABLES

(January, 1986, and June, 1986, Review Sample Data)

Variable 1 (Thought and Detail)

Reader No.	Reader No. 1	2	3	--------	37
1		r(1,2)	r(1,3)	---	r(1,37)
2	r(2,1)		r(2,3)	---	r(2,37)
3	r(3,1)	r(3,2)		---	r(3,37)
-	-	-	-		-
-	-	-	-		-
37	r(37,1)	r(37,2)	r(37,3)		-
Mean Correlation	$\bar{r}(1)$	$\bar{r}(2)$	$\bar{r}(3)$		$\bar{r}(37)$

Note: Each value in the table represented a correlation between the scores awarded by two readers on scoring variable 1. For example, r(3,2) was the correlation between readers 3 and 2 on the scores they awarded for scoring variable 1. As in Table III-4, all values appeared twice, and no entries were made where the reader numbers would be the same.

The table then continued, giving values for scoring variable number 2, 3, and so on to variable 11.

deviation for each paper.

Once the compositions having low scorer reliability were identified, they were categorized according to common characteristics. These characteristics could not be predicted before the data were analyzed; rather, they were determined after inspecting the problem papers.

B. Reader-Based Problems

The degree to which readers tended to submit discrepant scores was determined. Weighted scores awarded by the readers on the reliability sample compositions were again employed. These scores were used to calculate correlations between all pairs of readers. Next, the mean correlation for each reader was determined (see Table III-4). Readers associated with the lowest mean correlations were identified as being the most discrepant.

It must be noted that correlation coefficients reflected only one aspect of scoring discrepancies. They indicated the extent to which readers agreed in the ranking of the papers, but ignored differences in standards. For example, two readers could have a correlation of 1.0, indicating perfect agreement in ranking the essays, but one reader might award marks that were consistently ten points lower for every essay.

Discrepancies in standards were determined using another criterion. This criterion consisted of the means and standard deviations of the scores awarded by readers over all of the review sample compositions (see Table III-3). A low mean indicated that the reader was marking more severely than the others, and conversely, a high mean indicated a generous reader relative to the others. High or low values of the standard deviation indicated readers who scored over a relatively wide range or narrow range, respectively.

C. Problems Related to the Scoring Variables

The effectiveness of the various scoring descriptors was judged. Data from the combined review samples were used to determine the descriptors for which the reliability of scoring was comparatively low (see Table III-5). The combined samples were used because many readers (thirty-seven) were involved in both the January, 1986, and June, 1986, scoring sessions. While the writing topics given to the students differed, the scoring variables did not. For each variable, correlations between all pairs of raters were calculated. Variables for which the mean correlation was low were identified. Again, what was deemed to be low was determined after the data were examined.

D. Problems Related to the Writing Tasks

Writing tasks for which the reliability of scoring, in the review samples, was lower than others were identified. For each of the three writing assignments in an examination paper, the reliability of scoring was estimated by calculating the mean correlation over all pairs of readers. Comparisons revealed any writing task that was scored less reliably than the others.

II. Assessing Test Reliability

The reliability of the written test as a whole, as opposed to the reliability of the scoring, only, was calculated. Scoring reliability was concerned with discrepancies among the readers when each student submitted an essay. Test reliability, in addition, took into account differences in quality of the writing. For example, a student might write a high quality essay in one instance, and a lower quality essay in a second instance. An alpha coefficient based upon all the scoring variables was computed. This coefficient was a measure of internal consistency, similar to the well known KR20 estimate of reliability used in objective tests. The difference is that KR20 applies when test items are either right or wrong and are awarded a score of one or zero. An alpha coefficient applies when items could have a wide range of scores. The weighted scores actually awarded to the essays were used; these scores were based upon three independent readings.

The purpose of this procedure was to determine the effectiveness of the methods employed by Alberta Education to increase reliability. The value was expected to be higher than 0.8, (Diederich, 1974, p. 34; Akeju, 1972, p. 178) in view of the special measures taken.

The improvement in reliability that resulted from a consensus reached from three gradings instead of from a single scoring was calculated. The reliability of the test as a whole (coefficient alpha) was computed in connection with objective B. 1. (see page 6). The reliability (coefficient alpha) of the scores submitted after the first reading of the compositions from the random sample of ninety-six was computed and compared with the coefficient calculated from the final marks that were awarded. These final grades were based upon scores given by all three readers.

Summary

The problem of essays for which discrepant scores were submitted was addressed through data gathered from the twice-daily reliability review sessions. Problems related to the compositions were studied by comparing the standard deviations of the scores awarded by all of the raters. Reader-based problems were investigated, first, by determining which markers tended to have the lowest correlations with other markers; second, by identifying readers who tended to award scores that were too high or too low, as compared to the other readers; and third, by identifying readers whose score spreads were high or low. Problems related to the variables were identified by determining which ones were associated with low correlations amongst readers. Similarly, problems related to the writing tasks were identified by determining which tasks had low correlations amongst readers. The alpha coefficient, calculated from a random sample of papers, was taken as an estimate of test reliability. Comparisons between coefficients based upon the first reading of the essays and upon the final grades determined from three readings were made.

CHAPTER IV

OBSERVATIONS AND DISCUSSION OF RESULTS

I. Compositions Having Low Scorer Reliability

A. Problems Related to the Compositions

The proposed procedures (see page 51) for identifying problem papers were carried out. Standard deviations of scores given by all readers were calculated for each of the reliability review papers (see Table IV-1). Next, the papers for which the standard deviations were high (that is, when the reliability of scoring was low) were analyzed with a view to pinpointing the causes of the disagreement among scorers. A detailed analysis of the problem papers for the January, and June, 1986, scoring sessions was carried out. Following this analysis, generalizations were made regarding what constituted problems for many markers. In the study, inferences based upon subjective judgments often were made. These judgments seemed justified in view of the extensive experience of the author as a group leader during scoring sessions and in view of conferences held with other markers.

1. January, 1986, Reliability Review Papers

Statistical calculations revealed that during the January, 1986, scoring session, the standard deviations for the seven reliability review papers ranged from a high of 12.658 (paper five) to a low of 5.473 (paper three). The seven papers that were scored by all seventy-five readers were examined to determine the extent to which these examinations caused difficulties for the markers. These seven papers were studied in detail and the results were summarized in the following pages.

Table IV-1

MEANS AND STANDARD DEVIATIONS OF SCORES
ASSIGNED TO RELIABILITY REVIEW PAPERS

January, 1986, Session			June, 1986, Session		
Paper #	Mean	S.D.	Paper #	Mean	S.D.
5	73.840	12.658	9	52.397	9.614
2	49.507	8.427	6	60.968	9.244
4	65.973	8.426	2	54.953	8.757
7	58.982	7.802	8	73.141	8.387
1	45.973	7.793	7	77.156	8.339
6	65.068	7.336	1	62.797	8.120
3	90.680	5.473	4	72.638	8.106
			5	41.746	7.368
			3	85.206	7.236
			10	67.585	6.939

Paper Causing Greatest Problem - Jan. R.R. #5

Reliability review paper number five (see Appendix B Part 1) caused the largest problem for the markers (standard deviation of 12.658) and the difficulty had much to do with the writer's treatment of the subject in section one of the exam. The assigned topic (topics for January, and June, 1986, are found in Appendix C) was to "write about an incident that focuses on the giving or receiving of a special gift" (Alberta Education, 1986, p. 3). Students were asked to "describe the incident, examine the feelings and attitudes created by the incident, and explain how this incident has affected you" (Alberta Education, 1986, p. 3). In paper five, the student chose to write about an intangible gift, the gift people have of being talented in more than one area. It should be noted that common responses to the topic dealt with articles that students had received as gifts (females often wrote about receiving a piece of jewelry from boy friends; males commonly wrote about getting their first car). The writer of paper number five described three aspects of what was termed "giftedness"--talent in carpentry and fixing cars, and the person's ability to get along with others. The paper concluded by describing the personal satisfaction gained when others were helped. These ideas were not commonly found in most papers written by English 33 students and, in addition, they were expressed in a sincere way.

In section one of the examination, seventeen markers awarded the highest possible grade of five for variable one, thought and detail. One marker must have believed that the ideas did not deal with the prescribed topic and thus gave the lowest possible grade of INS (insufficient) and fifteen others gave grades of one or two for variable one, thought and detail. The marks awarded for the first scoring variable spread over the entire range of possible grades, and the paper appeared to cause markers problems because of the unusual treatment of the topic.

The halo effect seemed to enter into the scoring of paper five in that seven of the markers awarded the grade of five to the other three variables in section one (organization, matters of choice, matters of convention) and five of these same markers awarded the grade of five for the three variables contained in section two of the examination. It was somewhat difficult to defend the grade of five in the category of matters of convention ("The writing is essentially free from errors in spelling, punctuation, and grammar. Errors that are present do not reduce the clarity of communication" [Student Evaluation, 1986b, p. 33].) because there were numerous errors in all three of the cited areas. The halo effect also appeared to cause the marker who gave the grade of INS in section one to continue giving lower grades than the scoring descriptors called for in section two of the examination.

In section two, the students were asked to invite a speaker to the school to "represent the career or occupation that most interests you" (Alberta Education, 1986, p. 11). Students were told that the letter should demonstrate "your interest in the career or occupation that your speaker will represent" (Alberta Education, 1986, p. 11) and that it should include "all of the information that your speaker will need to make his or her participation successful" (Alberta Education, 1986, p. 11). A sample poster that had been placed on a school wall was included which provided more information for writers. Paper number five was one and one-half pages in length and was longer than most compositions written by English 33 students. The letter listed the purpose in writing, mentioned the paramedical field, cited questions that students would ask, and asked for an early reply.

In section two of the examination, the greatest differences in grades appeared to occur because of the halo effect continuing from section one. If, in the first writing task, markers had not followed the scoring descriptors and had awarded higher or lower grades than could be expected, they continued to do so in section two of the test.

Section three of the examination consisted of two questions based upon two cartoons. The first question asked students how they would feel if they were one of the characters in the cartoon. They were asked to provide support for their answers. The writer of paper number five gave a response that was a page in length; this was a longer answer than was typically given. The question was answered in a slightly different way in that feelings were not expressed in a direct manner. Instead, the writer used a rhetorical question about hitting a finger with a hammer.

Five markers awarded the grade of five for thought and detail while four markers gave the grade of one in the same category. The generalizations that had been expressed in the answer seemed to cause problems for the markers.

In question two of section three, students were asked to list the main idea that was common to both cartoons; they were then to support their response with details from the cartoons. The writer of paper five gave a lengthy answer that covered the activities shown in the cartoons. Although the statements were true, they did not deal with the main idea common to both drawings.

Seven markers awarded a grade of four for variable ten, thought and detail, even though the scoring guide for capable writing did not appear to fit well ("Interpretation of the comic strip, cartoon, or photograph is appropriate and is in the form of a generalized idea or theme. Specific details used for support are well-defined and accurate" [Student Evaluation, 1986b, p. 37].). Three of the markers who gave grades of two for variable ten, thought and detail, also gave the same grade to the writing skills section. It appeared that the scoring descriptors were not followed carefully, but that the same grade of two was arbitrarily given to the writing skills section.

Paper Causing Second Greatest Problem - Jan. R.R. #2

Reliability review paper number two (see Appendix B Part 2) caused the second largest problem for many markers (standard deviation of 8.427). Section one of this paper contained two diary entries. The first entry consisted of the story the writer had heard of a grandfather telling his grandson that life could be compared to the seasons. The second entry was brief and stated that the writer had the confidence to live a full life. This paper did not mention the word gift and, like paper five, contained a different idea than was typically found in the examination papers.

A number of markers appeared to have difficulty in grading variable one, thought and detail, because of the nature of the composition. Marks ranged from insufficient to four with the mean mark given being a grade of two. The two markers who gave the grade of INS seemed to have felt that the paper did not deal with the assigned topic; markers who awarded the grade of four might have thought that the content was creative and thus worth a higher grade. They might also have been influenced by the rather sentimental nature found in the treatment of the subject. Ten markers who gave the grade of two for thought and detail in section one also gave the same grade for organization ("The beginning and/or conclusion are non-functional. Relationships among sentences and between paragraphs are frequently unclear" [Student Evaluation, 1986b, p. 32].) It seemed that the halo effect entered into the scoring in these cases because the composition did contain a functional beginning, development, and conclusion and more closely fitted the adequate (grade of three) category. It should be noted that three markers who gave the grade of one for thought and detail moved to a two for organization. They believed the student-writer earned a lower grade for thought and detail and also penalized the writing for organization.

Section two of this student's paper contained a half-page letter inviting a speaker to talk about his occupation, although the specific occupation was never mentioned. Possible questions were cited and the writer concluded by stating interest in the speaker's occupation.

Grades for variable five, thought and detail, ranged from one to four. The grades of three and four seemed to be a little generous after the scoring descriptor was examined carefully in the marker's guidebook. The one marker who awarded the grade of four for thought and detail also awarded the same grade for organization in this section.

In section three question one, the writer did mention his feelings directly by stating that he would feel "discusted"; the response to the question was rather brief and was also stated in general rather than specific terms. These factors seemed to cause problems for some readers.

Marks for variable eight, thought and detail, ranged from one to four. Although there was no real problem with variable nine, writing skills, it should be noted that the mark of four that was awarded by two markers was somewhat questionable. The writer did not make many errors in this question; however, the response was very brief and there was little opportunity to make mistakes. Brief answers often caused problems because an accurate sample of the student's ability was not produced.

In question two of section three, the writer gave a main idea from the cartoons that was not particularly accurate. The writer stated that the idea behind the cartoons was that people could not afford fancy homes and that they had to build simple ones themselves. The marks given to this question ranged from one to four. The marker who awarded a four for variable ten had also awarded the same grade for variables eight and nine; perhaps the halo effect entered into the marking for variable ten. The grades of three and four seemed high if the descriptor were followed carefully. The ideas stated were not correct and the grade of two seemed to be more appropriate.

In the writing skills part of question two, the marks ranged from one to four. The marker who gave the grade of one also gave the same grade for variable ten, thought and detail; it appeared that the halo effect entered into the scoring of variable eleven.

Paper Causing Third Greatest Problem - Jan R.R. #4

Reliability review paper number four (see Appendix B Part 3) caused the third largest problem for many markers (standard deviation of 8.426). Section one of this paper dealt with working at a petting zoo for children. One little girl gave the worker a picture that included the words, "I love ya." The writer stated that the gift of a little girl's love had been received. Numerous mechanical errors were made throughout this composition.

The marks for variable one, thought and detail, in section one ranged from two to five. The content of the composition was mature and made direct reference to a gift that was received. Retrospection about the situation was directly stated. Three markers gave the grade of two for thought and detail and it seemed as though these individuals were penalizing the student for the number of writing errors that had been made throughout the composition. Matters of convention appeared to influence the grading for thought and detail.

The next problem in section one occurred with variable three, matters of choice. The marks ranged from two to five. One of the markers who awarded the grade of five for variable three had also awarded the same grade for the first two variables; it appeared that the halo effect caused the marker to pay little attention to the scoring descriptor definition.

The final problem in section one was found in the area of matters of convention, or variable four. Marks ranged from one to four for this variable. Although the writer made numerous spelling mistakes throughout the composition, the errors did not "severely impede the clarity of the communication" (Student Evaluation, 1986b, p. 33) and the grade of one seemed somewhat harsh. To award the grade of four meant that there were "few errors in spelling, punctuation, and grammar" (Student Evaluation, 1986b, p. 33), and this was not the case with the paper. It appeared that some markers were not following the scoring descriptors carefully when grading matters of convention.

Section two of the examination contained a lengthy letter which went on to the third page in the examination booklet. In the first paragraph, the writer requested the speaker to share some information with the school. In paragraph two, questions were asked about gemology, and in paragraph three, the writer asked the potential speaker to "show up" at career day.

The marks for variable five, thought and detail, ranged from two to five. The grade of five awarded by eight markers seemed to be somewhat high because the descriptor called for "precise and appropriate details" (Student Evaluation, 1986b, p. 34), and the writer gave details which were not particularly related to the assigned topic. It seemed likely that some of the markers were influenced by the number of details presented and by the length of the composition rather than by the quality of the work.

The marks ranged from two to five for variable six, organization. One of the markers awarded a five for thought and detail and also awarded a five for organization; it appeared that the halo effect was operating. Two other markers gave the grade of one for thought and detail and gave a two for organization; their lower grades also seemed to be influenced by the halo effect.

Variable seven, writing skills, had marks ranging from one to four. The single marker who awarded the grade of one for writing skills also gave the same grade for the conventions part of section one. This marker gave the grade of one for the writing skills part of section three, question one and appeared to be grading harder than other markers in the area of mechanics for this particular paper.

In section three question one, a lengthy one-page answer was given. The writer picked a single character from the cartoon and expressed many feelings about the situation in a direct manner. The feelings expressed were rather general in nature and much was inferred about the cartoon character.

Marks ranged from two to five for variable eight, thought and detail. Markers who awarded the grade of five might have been influenced by the length and general nature of the answer. The writer was creative in making up a situation that related to the character in the cartoon; however, the details were not "deliberately chosen for support and/or illustration of the feelings presented" (Student Evaluation, 1986b, p. 36).

Markers also had difficulty with variable nine, writing skills, and the marks ranged from one to five. One marker who gave the grade of one also gave the same grade for the category of matters of convention in other sections of the exam; the grader marked consistently harder for this paper. Another marker awarded the grade of five in other parts of the exam and awarded the same grade for variable eight; the marker graded consistently higher for this paper. Again, it appeared that the spread of scores was caused by markers failing to follow the scoring definitions.

In question two of section three, the writer expressed the idea that individuals wanted to have things that others had. The writer made a number of mechanical errors, but managed to express the theme behind the cartoon.

Markers had trouble with the category of writing skills. The five markers who awarded the grade of four awarded lower grades for the thought and detail section. It appeared that the markers reversed the categories, penalizing for thought and detail and rewarding for writing skills.

Paper Causing Fourth Greatest Problem - Jan. R.R. #7

Reliability review paper number seven (see Appendix B Part 4) caused the fourth largest problem for many markers (standard deviation of 7.802). Section one of this paper dealt with the story of a soapstone carver named Jonah. The boy gave his parents a carving he had created as a present and the couple hoped that their son would become a great artist. Retrospection was stated directly. Words were mixed up in this composition and errors were made in matters of choice and convention.

Only variable one, thought and detail, caused problems for markers in section one of the examination. The marks ranged from two to five and the problems that occurred might have been caused by the nature of the topic and by the errors that were made. The subjects discussed in the composition (a gift for parents, the development of talent in youth, and the sense of a boy's appreciation for what his parents have done for him) seemed to carry with them a form of sentimental attachment for certain markers. Some markers might have been influenced by the nature of the subject matter (either positively or negatively) and feelings rather than descriptors appear to have been followed. On the other hand, the grade given for thought and detail by some markers seemed to be affected by the number of errors in choice and convention that the writer had made. A number of graders seemed to be reluctant to separate matters of choice and convention from the category of thought and detail. Descriptors were not followed; instead, some markers seemed to grade on the basis of superficial feelings. One of the markers who awarded the grade of five for thought and detail chose to give the grade of four for matters of choice. The scoring descriptor was not followed, and it seemed likely that the halo effect entered into that marker's judgment about descriptor three.

Section two of the examination contained a letter that was just over a page in length. The first sentence of the composition was very long and the style was closer to that of a telephone conversation than a business letter. The occupation of R.C.M.P. was mentioned and justification for asking the speaker was given. The writer mentioned the need for pamphlets and slides in the presentation. Details were rather general in nature throughout the letter.

Marks for variable five, thought and detail, ranged from one to four. The single marker who gave the grade of one appeared to have had the problem of not following the complete wording for the descriptor. The final words of the descriptor were "the purpose is not fulfilled" (Student Evaluation, 1986b, p. 34), and these words did not apply to the letter. It seemed likely that the grader did not read the entire descriptor carefully before giving the grade.

In section three question one, the writer did not choose one of the characters in the cartoon directly but stated that pain would be felt by all the men in the drawing. The composition was a full page in length but many irrelevant details were cited.

Marks given for variable eight, thought and detail, ranged from two to five. It appeared that the three markers who awarded the grade of five were more influenced by the length of the answer than they were by its content. The scoring descriptor was not followed accurately by those markers who awarded the grade of five for thought and detail because the writer was to have assumed an appropriate role from the cartoon "effectively and consistently" (Student Evaluation, 1986b, p. 36); this requirement was not met.

In section three question two, the writer became a little confused about the main idea common to both cartoons but still wrote a full page response to the question. The student answered that some men do the same kind of work on the same type of building and that the six men were in pain. In another paragraph, the writer seemed to become even more muddled.

Marks for question two, variable ten (thought and detail) ranged from one to five. It appeared that the four marks of four and the one mark of five were awarded on the basis of the length of the composition and not upon the content of the answer. It also seemed likely that the answer was not read carefully and that the scoring descriptors were not well applied to the content of the composition. For variable eleven, writing skills, the marks ranged from one to four. The marker who gave the grade of one for writing skills also awarded the same mark for the category of thought and detail; it appeared that the halo effect carried over from variable ten to variable eleven.

Paper Causing Fifth Greatest Problem - Jan. R.R. #1

Reliability review paper number one (see Appendix B Part 5) caused the fifth largest problem for many markers (standard deviation of 7.793). Section one of this paper dealt with the story of a Christmas gift that had been given to a woman by her husband. The woman received a large box and opened it in front of her family. Inside the box was a bag of candy. When the upset woman reached into the bag, she found a diamond ring. The woman still remembered that her husband had said, "You always make a judgment before discerning or compairing." The writer attempted to utilize large words but they were often misspelled or misused. Judging by the content and style of the composition, it seemed likely that the woman was an adult student who had learned English as a second language.

Markers had trouble with variable one, thought and detail, because the grades ranged from two to five. The grade of two seemed somewhat harsh in that the retrospection was "conventional" (Student Evaluation, 1986b, p. 32) and not "obscure or uncertain" (Student Evaluation, 1986b, p. 32). The mark of two could have been given because of the problem the writer had with matters of choice and matters of convention; some markers penalized for thought and detail when they also penalized for mechanics. The grade of five that was awarded by one marker appeared to be generous when the retrospection part of the descriptor was followed; retrospection was not particularly "insightful" (Student Evaluation, 1986b, p. 32). Some of the markers might have been influenced either positively or negatively because the writer appeared to be an adult and an English as a second language student.

For variable two, organization, the marks ranged from one to four. The graders who gave the grade of one most likely confused the category of matters of choice with the category of organization. Some of the words in the introduction were used awkwardly or incorrectly, but the introduction and conclusion were present in the composition and certainly could not be considered to be "obscure" (Student Evaluation, 1986b, p. 32). It appeared that markers were penalizing for poor mechanics when they were supposed to be scoring for the category of organization. The two markers who gave the grade of one for organization, variable two, also gave the same grade for variables three and four (matters of choice and matters of convention), suggesting that the halo effect entered into the marking situation.

For variable three, matters of choice, the marks ranged from one to four. Three markers who awarded the grade of four might have been influenced by the writer's attempts to choose longer and more complex words. These words were chosen for effect, a requirement when awarding the grade of four; however, the writer did misuse many of the words. This misuse caused the grade of two to fit the composition better than the grade of four.

Section two of the examination consisted of a letter inviting a doctor to the career day function. The letter was a little more than a page in length and contained numerous mechanical errors. The style was rather awkward and was overly polite.

The marks ranged from one to four for variable five, thought and detail, in section two of the examination. The grade of four seemed to be rather high in that very little information was presented in the letter. It appeared that the three markers who awarded the grade of four were not following the descriptor carefully when grading for thought and detail; these markers might have been influenced by the very polite tone of the letter.

In section three, question one, the student mentioned that if she were Jack, she would feel stupid cutting the grass because it would be better to be inside. The wording was quite stilted and the composition was difficult to comprehend.

The marks ranged from one to four for variable eight, thought and detail. One marker awarded the grade of four and did not appear to apply the total descriptor definition to the answer written by the student. Although the writer did choose a character, she did not establish clear feelings and did not use "specific details" (Student Evaluation, 1986b, p.

36) for support that were "well-defined and accurate" (Student Evaluation, 1986b, p. 36). The writer expressed generalizations rather than specific details.

In section three, question two, the student wrote that people could be slaves to their houses but could also take things easily. It was also stated that people could rebel against their houses. The writer failed to grasp the message found in both cartoons, but did attempt to answer the question.

Marks for variable ten, thought and detail, ranged from insufficient (INS) to four. For the two markers who awarded the grade of INS, they must have believed that the student did not respond to the question that was asked. This grade seemed to be rather harsh in that the student did respond even though the answer was not particularly accurate. Markers often had trouble with the category of insufficient in that it was difficult to determine whether a divergent answer fitted the descriptor. The grade of four that was awarded by one marker seemed high in that the interpretation of the cartoon was supposed to be "appropriate" (Student Evaluation, 1986b, p. 37). It appeared that the descriptor was not followed carefully before awarding the grade.

For the final variable, writing skills, one marker awarded the grade of four and must have felt that "the selection and use of words and structures is often effective" (Student Evaluation, 1986b, p. 37). If the composition were read carefully, the marker should notice the large number of errors in the words that the writer used. The descriptor did not appear to fit the composition well and was likely chosen because of excessive speed in marking or because of careless grading.

Paper Causing Sixth Greatest Problem - Jan. R.R. #6

Reliability review paper number six (see Appendix B Part 6) caused the sixth largest problem for many markers (standard deviation of 7.336). Section one of this paper dealt with the story of two sisters who told each other what their parents were getting them for Christmas. The composition related an incident clearly; few errors were made, but the retrospection about the incident was stated rather briefly.

Markers had some difficulty with variable one, thought and detail, in that the marks ranged from two to five. The three markers who gave the grade of two likely believed that the grade was justified because they might have felt that the "retrospection or circumspection is obscure or uncertain" (Student Evaluation, 1986b, p. 32). The remainder of the descriptor for the grade of two did not fit the composition well and the markers likely had to make a decision whether to give a grade of two or three. Although markers were told to "waffle up" when in doubt, the three markers who gave the lowest grades appeared to have done the opposite. The two markers who awarded the grade of five did not follow the descriptor very effectively in giving the highest grade to this composition. Retrospection was not particularly "insightful" (Student Evaluation, 1986b, p. 32) and the markers should have noticed this before awarding the grade.

Marks for variable two, organization, caused a problem for one marker as evidenced by the grade range from two to five. The single marker who gave the grade of two had also given the same grade for the first variable. It was assumed that the halo effect was operating for the grader because the paper did not fit the descriptor listed for the grade of two.

Section two of the examination contained a letter which began by giving an introduction to Career Day and then went on to give an invitation to speak. No specific career was mentioned and the date was cited as January rather than June.

Variable five, thought and detail, in section two of the examination caused problems for a few markers. The marks ranged from one to four and three markers gave the grade of one in this section. These markers might have been influenced by the treatment of the subject. The assignment was to invite a speaker who represented a career that was of interest to the student. Markers could have felt that the treatment of the subject was different from the assigned topic in that no specific career was described. Markers might also have given the grade of one because of the change in the date from that listed on the assignment page.

Marks ranged from one to five for variable six, organization. The two markers who gave the grade of one had also given either the grade of one or two for variable five, thought and detail. The halo effect appeared to enter into the grading for these individuals and it seemed that they did not follow the descriptor carefully. The single marker who awarded the grade of five for organization might have read the descriptor quickly and decided that it fitted the composition. It did appear to fit some of the letter in that there was an "effective statement of topic or function" (Student Evaluation, 1986b, p. 34), but when the composition was examined carefully, it could not be said that it fitted the entire descriptor effectively.

Marks ranged from two to five for variable seven, writing skills. In both the high extreme and the low extreme cases, it appeared that the markers were not examining the composition carefully and were not applying what they had read to the scoring descriptors. It seemed difficult to see how four markers could say that the selection of words was "usually effective" (Student Evaluation, 1986b, p. 35) while thirteen others found those same words to be "often ineffective" (Student Evaluation, 1986b, p. 35).

In section three, question one, a six line response was given. The writer stated that she would feel "embarresed" being a carpenter and hitting her thumb.

Marks for thought and detail ranged from one to four. Both extremes of the marks awarded appeared to be caused by markers who were following part of a descriptor but not the entire definition. The writer did assume a "plausible role" (Student Evaluation, 1986b, p. 36) in the composition; if the writer had not, then the grade of one should have been awarded. The composition did not provide specific details that were "well-defined and accurate" (Student Evaluation, 1986b, p. 36); if it had, then it should have received the grade of four. It appeared that only part of the descriptor was followed by the discrepant markers.

In section three, question two, the writer stated that the main idea of the two cartoons was home maintenance; the response was just over ten lines in length.

Marks for variable ten, thought and detail, ranged from one to four. The six graders who awarded the mark of four did not appear to have followed the descriptor because the answer given was supposed to be appropriate, but this response had little to do with the meaning of the cartoons.

Marks for variable eleven, writing skills, ranged from two to five. The one marker who awarded the grade of five also awarded the grade of five for each of the other writing skills sections of the examination. The marker might have felt she was being consistent in awarding the same grade to the same student for the same category throughout the examination.

Paper Causing Seventh Greatest Problem - Jan. R.R. #3

Reliability review paper number three (see Appendix B Part 7) caused the least difficulty of the seven papers for the markers (standard deviation of 5.473). This paper was written effectively and utilized a mature style; compositions that fit the scoring descriptors at the top level ("five" papers) rarely cause markers a great deal of difficulty in scoring consistently.

Section one of this paper contained an account of the writer's failure to find his mother a birthday present. The writer recounted details from his past and stated that no event had moved him so much in his life. Sixty-one markers awarded the grade of five for variable one (thought and detail) in section one and twelve more awarded the mark of four. Two markers gave the grade of three and these individuals might have believed that the student's retrospection was brief and therefore was conventional. The effective situation had been established and the implicit aspects of the retrospection appeared to have been missed by the two markers.

One marker gave the grade of three for variable two, organization. This individual had given the same grade for variable one and it appeared that the halo effect was operating.

The letter written for section two of the examination contained information necessary for a speaker, but did not give much detail about the speaker's task. The letter was written in a mature style and was virtually error free.

Marks ranged from two to five for variable five, thought and detail. The five markers who awarded the grade of five had all awarded the same grade for each variable in section one of the examination. The halo effect seemed to carry over into this part of the examination.

For variable six, organization, the marks ranged from two to five. The two markers who gave the score of two and the eight graders who gave the mark of three might have been influenced by the three-quarter page length of the letter. They could have felt that the letter's brevity did not allow for details to be related to the topic. It should be noted that these two markers had also given a two for variable five; the halo effect might have entered into the scoring for this descriptor.

Two markers gave the grade of three for variable seven, writing skills. These individuals had given lower grades for all parts of section two; the halo effect appeared to be operating.

In question one of section three, the student wrote that he would try not to feel anything because everything was too much the same. This response was very different from others that had been written; the paragraph contained few errors.

The marks for variables eight and nine (thought and detail and writing skills) ranged from three to five. Two markers gave the grade of three for variable eight; these people might have been confused by the different nature of the response. The single grader who gave the mark of three for variable nine also gave the same grade for all parts of section three; the halo effect seemed to influence this individual.

For question two of section three, the student wrote that people were influenced by things around them; evidence from both cartoons was cited.

The marks for variable ten, thought and detail, ranged from two to five. One marker gave the grade of two and seven gave the grade of three. These scorers might have been influenced by the relatively brief answer that was given.

Variable eleven, writing skills, had marks which ranged from three to five. A single marker gave the grade of three, the same grade that he gave for all parts of section three.

2. June, 1986, Reliability Review Papers

Statistical calculations revealed that during the June, 1986, scoring session, the standard deviations for the ten reliability review papers ranged from a high of 9.614 (paper nine) to a low of 6.939 (paper ten). All of the papers scored by the sixty-four markers were studied in detail and the results were summarized in the following pages.

Paper Causing Greatest Problem - June R.R. #9

Reliability review paper number nine (see Appendix B Part 8) caused the largest problem for many markers (standard deviation of 9.614). The assigned topic was to write about a person who "has revealed an unexpected side of himself or herself" (Alberta Education, 1986b, p. 2). Students were asked to:

> Describe the person's public reputation, relate an incident that reveals an unexpected side of the person, examine feelings and attitudes resulting from the person's unexpected behavior, and explain why the person's unexpected behavior is memorable to you (Alberta Education, 1986b, p. 2).

In paper nine, the student wrote about a girl who was able to make others feel good about themselves. The writer used complex words and the composition contained few errors.

The marks for section one, variable one, thought and detail, ranged from one to four. The twelve markers who gave the grade of one might have had some problem with the topic that the student chose. It was not stated directly what the girl's unexpected side really was; however, it could be inferred that her positive attitude was different from the typical one of many people. These twelve markers must have felt that an "inappropriate situation" (Student Evaluation, 1986c, p. 30) had been presented. The two markers who awarded a grade of four for thought and detail did not seem to be following the scoring descriptor carefully in that the criteria of "specific details about actions and/or characters are well-defined and plausible" (Student Evaluation, 1986c, p. 30) did not fit this composition well. Generalizations rather than specific details were presented.

For variable three, matters of choice, the marks ranged from one to five. The two markers who gave the grade of one likely felt that the student's attempt to use complex words caused some difficulties. Some of these words were misused and this might have caused the markers to grade more harshly. In giving the grade of one, it seemed that personal bias entered into the grading. The errors in choices made by the student could not be considered to be "usually ineffective" (Student Evaluation, 1986c, p. 31). The single marker who awarded the grade of five likely felt that many of the words had been chosen for effect, but ignored the problem that some of these words had been used incorrectly. It appeared that the descriptor was not followed with care.

For variable four, matters of conventions, the marks ranged from one to five. In both the high and the low extremes, it seemed likely that markers were not reading the composition carefully when comparing it to the descriptors. There were neither "numerous errors in spelling, punctuation and grammar" (Student Evaluation, 1986c, p. 31), nor was the writing "essentially free from errors in spelling, punctuation and grammar" (Student Evaluation, 1986c, p. 31). Because this reliability review paper was scored near the end of the marking session, some graders might have felt that they had "internalized" the scoring descriptors and these people would likely not have taken the time to reexamine the marker's guidebook while grading this composition.

In section two, the students were asked to write a letter to the editors of a newspaper that identified "one problem that is of concern in your community" (Alberta Education, 1986b, p. 11); a list of possible problems was provided. Students were told to "describe, in detail, the problem that is of concern to you and tell what should be done about the problem you have described" (Alberta Education, 1986b, p. 11). Paper nine contained an answer that was more than two pages in length. The student wrote about the lack of community services for senior citizens from the point of view of a senior citizen. The composition contained generalizations rather than specifics about senior citizens and the student attempted to employ complex diction.

Marks for variable five, thought and detail, ranged from one to four. The three graders who awarded the mark of four did not appear to follow the scoring descriptor. Details which were provided in the composition were not "appropriate" (Student Evaluation, 1986c, p. 32) and did not "efficiently fulfil the purpose" (Student Evaluation, 1986c, p. 32). The problems that some individuals had might have been caused by markers who did not examine a descriptor carefully because of "internalized" standards. Variable six, organization, had marks which ranged from one to five. The halo effect seemed to be operating for the markers who awarded both the high and the low grades. The marker who gave the grade of one for organization also gave consistently low grades for the entire paper, and the other grader who awarded the five awarded consistently high grades for the paper. Feelings rather than descriptors were likely followed.

The grades for variable seven, writing skills, caused difficulties because the marks ranged from one to four. Markers who gave the grade of one might have penalized the composition because of the student's attempts to use larger words. The descriptor did not appear to have been followed carefully in that the errors in spelling, punctuation, and grammar did not "often impede the clarity of communication" (Student Evaluation, 1986c, p. 33).

Section three of the examination consisted of two questions based upon one cartoon. The first question asked the student to choose one of the characters in the cartoon, describe that character's feelings, and support the answer with details from the cartoon. The second question asked for the main idea that was communicated by the cartoon; details from the cartoon were to be used to support the answer that was given.

Variables eight and ten caused problems for markers. Both of these variables consisted of the category of thought and detail for questions one and two respectively. In the first question, the student gave a one page answer and stated directly that the girl in the cartoon was "feeling a sense of achievement." Few errors in mechanics were made in this composition.

The response moved away from the assigned topic in the final sentences; however, the student expressed a very plausible emotion. The emotion was not supported with a great deal of evidence from the cartoon and probably should not have been given the grade of five. The marker who awarded the grade of five might have been influenced by the length of the answer. The two markers who gave the grade of one likely felt that the writer strayed off the topic and thus gave the lower grade; however, the descriptor was not followed well and it seemed apparent that feelings entered into the scoring for this variable.

In answering the second question, the student stated that the main idea was that children had different views on life. Variable ten, for which scores ranged from one to five, was obviously a problem for markers. The single marker who gave the grade of one had also given the same grade for variable eight. The halo effect appeared to enter into the marking. The marker who awarded the grade of five might have agreed with what the writer had said about life and thus awarded the generous mark; however, the answer given did not appear to fit the five descriptor for thought and detail.

Paper Causing Second Greatest Problem - June R.R. #6

Reliability review paper number six (see Appendix B Part 9) caused the second largest problem for many markers (standard deviation of 9.244). Section one of this paper was written in the form of a letter to the local newspaper. The writer described a hockey player who was calm in life yet who was aggressive when he played hockey.

Marks for variable one, thought and detail, ranged from one to four. The composition was rather different in that it took the form of a letter and that form did not seem particularly appropriate to the topic; it might have been chosen by the student as a simple way to begin and end a piece of writing. The content of the writing was somewhat garbled in nature and the message was not totally clear; these factors might have caused the two markers who gave the grade of one to follow their inclinations rather than the descriptor definition.

Marks for variable two, organization, ranged from one to five. The single grader who awarded the score of five graded consistently high for all categories in sections one and two of the examination; the halo effect appeared to influence the grading of the paper. The grader who gave the mark of one for variable two had also given a low grade for variable one. This marker might have been influenced by the halo effect. In both the high and the low cases, it appeared that the descriptors were not followed with care.

Marks for variable three, matters of choice, ranged from two to five. The single marker who awarded the score of five was the same individual who created the problems with variable two by awarding a generous mark.

For variable four, matters of convention, the marks ranged from two to five. For the three markers who gave the grade of two and for the four markers who awarded the grade of five, it appeared that they were not able to follow the descriptors with accuracy. If the composition were read carefully, neither the highest nor the lowest mark that was awarded fitted the piece of writing. Marks seemed to be awarded without care and attention.

In section two of the examination, the student discussed the condition of the town community hall and it was suggested that all the members of the community could rebuild the hall. The letter was approximately two pages in length and was written in a clear style.

For variables five (thought and detail), six (organization), and seven (writing skills), the marks ranged from two to five. The grade of two that was given by some markers for each of the three variables in section two was a grade that did not particularly fit the composition. It appeared that the markers who awarded the grade of two did not follow descriptor definitions carefully.

In section three, question one, the writer stated that the children were playing "shurades" and that one boy was going to perform the part of a plane. The final paragraph of the composition contained some muddled wordings and mechanical errors, and was therefore rather confusing.

Marks for variable eight, thought and detail, ranged from INS to four. The writer picked a different character from the girl who was usually chosen and the student expressed feelings that were somewhat hidden within the response. For one marker, this might have been enough to make him believe that the composition did not respond to the question that was asked and the grade of INS was given. It appeared that the marker who awarded the grade of four did not follow the descriptor for thought and detail. The emotion experienced by one of the characters was not "clearly described" (Student Evaluation, 1986c, p. 34) nor was it "appropriate" (Student Evaluation, 1986c, p. 34). The details used for support were also not "well-defined and accurate" (Student Evaluation, 1986c, p. 34). This marker had awarded the grade of four for variables two through nine and it appeared that the grade was assigned because of the halo effect.

In section three, question two, the writer stated that the main idea of the cartoon was that people enjoyed themselves at school because teachers were nice.

The marks ranged from one to four for variable ten. The single marker who awarded the grade of four and the ten markers who gave the grade of three might have been influenced by the subject. The ideas stated presumably were pleasing ones for teachers although they had little to do with the meaning of the cartoon. The nature of the subject matter might have caused the markers to award grades that were higher than what following the descriptor would have warranted.

Paper Causing Third Greatest Problem - June R.R. #2

Reliability review paper number two (see Appendix B Part 10) caused the third largest problem for many markers (standard deviation of 8.757). Section one of this paper dealt with the story of Paul Lloyd, the "king" of Aspen playground. The younger children were afraid of him, but the older group saw "threw" him. An example was provided from the summer before. The composition was chronologically inconsistent and contained many mechanical errors.

The marks for variable one, thought and detail, ranged from one to four. The single grader who gave the grade of one also gave six more grades of one in other categories of this paper. Either the halo effect entered into the marking procedure or the individual was grading consistently hard.

Marks for variable three, matters of choice, ranged from two to five. One grader awarded the score of five for variable three; this person might have been influenced by the writer calling the boy a "rabid dog, ready to lash out" or by the slightly muddled words from the song, "he was bad leroy Brown the baddest dude in the hole town;" however, these words that were chosen for effect did not fit the descriptor for the grade of five. "Many words are chosen for effect and most are correctly used" (Student Evaluation, 1986c, p. 31) simply did not fit this composition better than the descriptors given for either the grades of three or four.

For variable four, matters of convention, the marks ranged from one to four. The two graders who awarded the score of four might not have read the composition carefully before comparing it to the scoring definition. The student's handwriting made the piece somewhat difficult to read and if the markers read the paper quickly, they might have missed the mechanical errors that were made throughout the writing.

In section two of the examination, a two-page letter was written outlining the problems of snow removal and road maintenance in the fictional town of Nalwen.

Marks for variable five, thought and detail, ranged from one to five. The single marker who gave the grade of one had also given the same score for variables one and four and for all of the parts of section two of this examination. This marker might have disagreed with the somewhat flippant tone used in this composition and could have scored the piece based upon his feelings rather than by following the scoring definitions listed in the guidebook.

Marks for variable six, organization, ranged from one to four. The grader who gave the mark of one also gave the same grade for all parts of section two (see paragraph above).

Variable seven, writing skills, had marks ranging from one to four. The one grader who awarded the mark of four also awarded higher grades for all parts of section two and the halo effect appeared to affect the scores.

In section three, question one, the writer stated that the boy thinking about the sailboat was "probaly wonding how he is goiing to do it" and that he did not think he could act out the part. The response was brief and was difficult to follow because of the errors made in choices and conventions.

Marks ranged from INS to three for variable eight, thought and detail. The single marker who gave the grade of INS might have been somewhat confused because she did not give the same grade for variable nine. When the grade of INS was given, markers were required to award the same grade for all parts of that section of the examination. It seemed likely that the marker was attempting to award the lowest possible grade and chose the grade of INS rather than the more appropriate grade of one. The marker who awarded the grade of four might not have followed the scoring descriptor carefully; this grader awarded the score of four for variables nine, ten, and eleven.

For section three, question two, the student wrote that the cartoon showed the differences among people and that people thought differently. The half-page response contained numerous errors in mechanics.

Variable eleven, writing skills, caused problems for the marker who awarded the grade of four on variables nine, ten, and eleven. It appeared that the scoring descriptors were not followed.

Paper Causing Fourth Greatest Problem - June R.R. #8

Reliability review paper number eight (see Appendix B Part 11) caused the fourth largest problem for many markers standard deviation of 8.387). Section one of this paper contained the story of a young man from a troubled home who turned to crime following the death of his dog. Few mechanical errors were made in this two-page composition.

The marks for variables one and two, thought and detail and organization, ranged from two to five. A single grader gave the mark of two for both variables. This marker might have been negatively influenced by the sentimental nature of the subject and thus awarded lower grades than what the descriptors called for. Personal bias appeared to enter into the scoring.

For variables three and four, matters of choice and matters of convention, the marks ranged from two to five. The grade of two seemed to be rather harsh when the descriptor definitions were compared to the composition. One marker gave the grade of two for both variables while two others gave a two for matters of convention only.

In section two of the examination the road conditions in and around Nalwen were outlined. The letter was written rather clearly and contained few mechanical errors.

The marks ranged from two to five for variables five, six, and seven (thought and detail, organization, and writing skills). One person gave the grade of two for variables five and six while another marker gave the same low mark for variable seven. This composition did not fit the mark of two that was given for all three variables; it appeared that descriptors were not read carefully before the mark was given in each case.

In section three, question one of the examination, the student stated that the girl probably felt that she was getting her idea across, when she actually was not. The writer repeated ideas throughout the composition; however, few mechanical errors were made.

Marks for variable eight, thought and detail, ranged from one to four. The six graders who gave the score of one and the seven graders who awarded the grade of four might have been a little careless in giving the low and the high scores. The definitions provided in the marker's manual did not appear to fit either extreme very effectively.

In section three question two, a half-page response was given stating that the main idea in the cartoon was about things people do and about how others see those things.

Marks for variable ten, thought and detail, ranged from two to five. The four graders who awarded the grade of five might not have followed the entire scoring definition carefully. The interpretation of the cartoon was not particularly "insightful" (Student Evaluation, 1986c, p. 35) and the mark of five appeared to be rather generous. Markers might have given a higher grade because the answer was more accurate than what commonly was found as the answer to the question.

For variable eleven, writing skills, the marks ranged from two to five. One grader gave the score of two for this section. The mark of two seemed to be a little harsh in view of the descriptor; however, the marker who gave the two had also given the same grade for variables seven, eight, and ten. This suggested that the halo effect was operating for variable eleven.

Paper Causing Fifth Greatest Problem - June R.R. #7

Reliability review paper number seven (see Appendix B Part 12) caused the fifth largest problem for many markers (standard deviation of 8.339). Section one of this paper dealt with the story of a ringleader of the underworld who was cruel to his enemies but kind to his family. The composition was a little different from many others in that it narrated a story and did not directly state the two sides of the man's character. Retrospection was not given directly but was implied within the story.

In section one of the examination, the marks for variable one, thought and detail, ranged from two to five. The seven graders who gave the score of two might have had trouble with the retrospection part of the composition because it was implied. This might have made the retrospection appear to be "obscure or uncertain" (Student

Evaluation, 1986c, p. 30). The remainder of the descriptor did not fit the composition very effectively and the markers would have been more consistent if all of the descriptor had been followed.

In section two of the examination, the letter to the editor discussed the lack of recreational facilities for young people but the writer did not deal with the solution to the community's problem.

The marks for variable four, thought and detail, ranged ~~from~~ two to five. The score of five seemed somewhat generous in light of the fact that the details given in the composition did not appear to be "precise and appropriate" (Student Evaluation, 1986c, p. 32). The student wrote about a concern that was of interest to youths and adults and the treatment of the subject might have influenced the markers in awarding more generous grades.

Variable five, organization, had marks which ranged from two to five. Three of the four graders who gave the score of two for variable five had also given the same score for variable four. The halo effect appeared to enter into the marking of this variable and the descriptor might not have been followed with care.

In section three, question one, the student wrote that the girl was trying to communicate the idea of a flower and was ignoring the rest of the class. The answer given was just over four lines in length.

Marks ranged from one to five for variable eight, thought and detail. The short answer might have caused the single marker to give the grade of one; however, the three markers who awarded the grade of five likely were not following the scoring descriptor accurately. The "precise details" (Student Evaluation, 1986c, p. 34) needed in the response to receive a five did not appear to be present in the answer written by this student.

Marks for variable nine, writing skills, ranged from two to five. Markers might have had some problem with this category because of the brevity of the response. Few mechanical errors were made in the composition, but the student had little chance to make errors in writing a four line answer.

In section three, question two, the student wrote that the main idea involved the differences in interpretation that people received. The response summarized the meaning clearly, but did not emphasize the details from the cartoon.

Marks for variable ten, thought and detail, ranged from two to five. The single grader who gave the score of two was the same individual who had given the score of one for thought and detail in variable eight. This person appeared to be marking consistently harder for the two variables on this paper, and it did not appear that the marker was following the scoring descriptors accurately.

Paper Causing Sixth Greatest Problem - June R.R. #1

Reliability review paper number one (see Appendix B Part 13) caused the sixth largest problem for many markers (standard deviation of 8.120). Section one of this paper contained a letter which described how a friend had repaired an object. The writer had not realized that his friend had any mechanical abilities. The handwriting caused this composition to be quite difficult to read; a number of errors were made throughout the piece.

Marks for variable one, thought and detail, ranged from INS to four. One grader gave the score of INS and might have felt that the composition did not respond to the assignment as stated. Perhaps the marker was expecting a personality change in an individual rather than the demonstration of an unexpected talent. Another grader gave the score of one, a grade that seemed to be rather low when the composition was examined carefully. The definition for the grade of one called for "an inappropriate or incomprehensible situation" (Student Evaluation, 1986c, p. 30); this composition had established rather a clear situation. It appeared that the descriptor was not followed carefully.

For variable two, organization, the grades ranged from two to five. The composition followed a logical pattern from beginning to end and the single grader who chose the grade of two might not have followed the descriptor carefully before giving the lower score.

For variables three and four, matters of choice and matters of convention, the scores ranged from two to five. The markers who awarded the grade of five for both variables might not have examined the composition carefully while they were matching it to the definitions. It appeared that the choices were not particularly proficient and that the composition was not "essentially free from errors" (Student Evaluation, 1986c, p. 31). It was possible that the student's handwriting caused some markers to miss the errors that were made.

In section two of the examination, a letter was written describing the poor condition of the main road through Nalwen. The problems cited seemed to be a little disjointed in nature and the composition did not flow well.

Marks for variable five, thought and detail, ranged from two to five. Two graders awarded the grade of five and these individuals must have felt that "significant information is presented, and this information is enhanced by precise and appropriate details that effectively fulfil the purpose" (Student Evaluation, 1986c, p. 32). The composition contained rather a simplistic solution to the road problem and the writer appeared to believe that none of the additional work would cost more money. The markers might not have looked at the content of the writing carefully before awarding the highest possible grade.

Marks for variable seven, writing skills, ranged from one to four. The single grader who gave the grade of one had also awarded lower grades for the matters of convention variable and for all the writing skills variables on this paper. This marker must have felt that the writer was consistently weak in the mechanics area.

In section three, question one, the student wrote that the boy who was thinking about the airplane was not interested in the girl's dance. Marks for variable eight, thought and detail, ranged from one to five. Two graders gave the score of one, and one grader gave the score of five. Both of these extremes seemed to be rather discrepant when the scoring descriptors were matched to the composition.

Variable nine, writing skills, had scores that ranged from two to five. The single marker who gave the grade of two had also given a lower score for variable seven, the writing skills variable in section two of the examination. It was possible that the halo effect became a factor because this grader had also given the same grade for variable ten.

In section three, question two, the writer stated that the main idea was that "growing children have a big imaginations." The response was just over five lines in length and contained a number of mechanical errors.

Marks for variable eleven, writing skills, ranged from two to five. The two markers who awarded the grade of five for the writing skills variable in question two had also awarded the same grade for the writing skills variable in question one. The halo effect might have influenced the scoring of variable eleven because the answer was quite brief and yet it contained a number of mechanical errors. Through careful following of the scoring definition, it would be difficult to award the grade of five in this case.

Paper Causing Seventh Greatest Difficulty - June R.R. #4

Reliability review paper number four (see Appendix B Part 14) caused the seventh largest problem for many markers (standard deviation of 8.106). Section one of this paper contained a letter which discussed a young man who appeared to be happy and successful but who had committed suicide. The composition was similar to the poem "Richard Cory" and the writer had adopted basically the same point of view.

In section one of the examination, the marks for variable one, thought and detail, ranged from two to five. The four graders who gave the score of two might have been troubled by the similarity between this composition and E. A. Robinson's poem. The descriptor definition for the grade of two did not fit this composition well and it appeared that another factor might have affected the scoring. That factor could have been a feeling that the subject was not original and that the markers had heard the material before.

Section two of the examination contained a letter written by a businessman who stated that he had moved to one of the "richer areas in town" but that his garbage was not being collected. The writer suggested that people should save their garbage and dump it on the lawn of city hall. The letter expressed a different situation than what most writers wrote about and the course of action that was described was quite original.

Marks for variable five, thought and detail, ranged from two to five. The three graders who gave the score of two might have been upset with the writer's subject and tone in the composition. The writer described himself as a wealthy man, discussed a situation that likely would not have occurred, and concluded with a scheme which threatened city hall. These factors might have caused the markers to give a lower grade for variable five.

Variable seven, writing skills, had marks that ranged from two to five. The one grader who gave the score of two could have been influenced by the rather sarcastic words that the writer chose in his composition. Although errors were made in this piece, their frequency did not seem to warrant being given the grade of two. It seemed likely that the writer's tone entered into the giving of the mark of two for one marker.

In section three, question one, the writer used the first person point of view. The writer stated that he wanted to leave the boring classroom to go fishing but that he was afraid he would be caught by the teacher again. This paragraph was different from most in that the feelings were expressed as if the writer were a character in the cartoon. Some errors in mechanics were made.

Marks for variable eight ranged from one to five. The graders who gave the score of one and two might have been troubled by the different tone the writer chose. The writer displayed a rather negative attitude toward school and teachers and this might have influenced some of the markers. The writer's idea about fishing was not a typical response and some markers might not have been able to accept a different answer without penalizing it.

In section three, question two, the writer stated that everyone was different with different "ideas, wants, and worries." It was mentioned that the cartoon also showed that life was a "wondrous and grand thing." The response moved from the message of the cartoon to a rather grand statement of theme.

Marks for variable ten, thought and detail, ranged from two to five. The four graders who awarded the score of five might have been influenced by the rather sweeping theme that the student stated. The theme did move away from the meaning found in the cartoon; however, some individuals might have liked and rewarded the general idea that was expressed.

Paper Causing Eighth Greatest Difficulty - June R.R. #5

Reliability review paper number five (see Appendix B Part 15) caused the eighth largest problem for many markers (standard deviation of 7.368). The composition written for section one of the examination dealt with the story of a boy who lied and got another boy into trouble. The writer concluded by stating that she wondered what happened to the boy and that she hoped that he "changed to good person." The writer appeared to be an English as a second language student; many mechanical errors were made. Marks for variable one, thought and detail, ranged from one to four. Five graders gave the score of one and these individuals might have been influenced by the nature of the composition.

The writer appeared to be an E.S.L. student and some of the sentences were rather stilted and difficult to understand. If markers became frustrated with the writing style and mechanics of the composition, they might easily penalize in the category of thought and detail.

Marks for variable two, organization, ranged from one to four. The single grader who gave the score of one for variable two had also given the same grade for variables one and three. It appeared that the halo effect entered into the scoring procedure.

The next problem with the scoring of this paper occurred in section three of the examination. The student wrote that she chose the character who was thinking of a tree. The character thought of a tree because the girl's hands were raised up, her body was skinny, and her legs looked like the "stem." The same problem with mechanical errors existed in this part of the examination.

Marks for variable eight, thought and detail, ranged from INS to five. Two individuals gave the grade of INS, four others gave the grade of four, and one person awarded the grade of five. All of these markers giving marks at the extremes appeared to be having difficulty applying the scoring descriptor definitions to a paper that was written by an E.S.L. student. The message should not have been considered to be insufficient, nor should it have been considered capable or proficient; however, when the message became muddled with the awkward wording, some scorers seemed to have had difficulty in determining the most appropriate grade.

Section three, question two, contained a difficult message to understand. The student wrote that the message was that "people act that it should be what it like while she was acting but people as the audience thinks it was difference way." The message was somewhat garbled and many mechanical errors were made in this paragraph.

Marks for variable ten ranged from one to four. Five graders gave the mark of one and these people might have been influenced by the confusing beginning of the composition. It was difficult to understand the student's message in the first sentence, but later the meaning became more clear.

Paper Causing Ninth Greatest Difficulty - June R.R. #3

Reliability review paper number three (see Appendix B Part 16) caused the ninth largest problem for many markers (standard deviation of 7.236). Section two of this paper contained the first problem for scorers. The writer discussed the need for good day care facilities so that mothers could go back to work. The composition was one and one-half pages in length and contained a few mechanical errors.

Marks for variable five, thought and detail, ranged from two to five. The single grader who gave the score of two might have felt that the "supporting details are scant and haphazard" (Student Evaluation, 1986c, p. 32) because the letter was somewhat shorter in length than many of the other compositions that had been marked. It might also be possible that the grader did not agree with the subject matter of day care and thus penalized the writer for the choice of subject.

Section three, question one of the examination described a little girl who was sharing her feelings about the music that was playing. Evidence from the cartoon was given directly which supported the writer's views.

The marks for variable eight, thought and detail, ranged from two to five. One grader gave the score of two. This individual might not have believed that a feeling was stated because the feeling was discussed later in the response. It was possible that the statement of feeling was skimmed over or missed by the marker.

In section three, question two, the student wrote that people had different ideas "even from an early age." It was mentioned that the girl was attempting to communicate the image of a flower but that each of the other class members perceived different pictures. It was also mentioned that each student had a different picture and that "these different pictures illustrate the different ideas each child can have." The word "different" was repeated several times in the last two sentences.

Marks for variable ten, thought and detail, ranged from two to five. The single grader who gave the score of two could have been troubled over the repetition of the wording about different ideas. Because of this repetition, it might have appeared that the student had not answered the question completely but that he was merely repeating ideas again and again.

Paper Causing Tenth Greatest Difficulty - June R.R. #10

Reliability review paper number ten (see Appendix B Part 17) caused the least difficulty of the ten papers for the markers (standard deviation of 6.939). Section one of the examination contained the story of a young man who revealed his feelings for his girlfriend after he was hurt during a hockey game. Retrospection was stated directly.

Marks for variable one, thought and detail, ranged from two to five. Three graders gave the score of two. When the descriptor definition was examined, the composition did not appear to fit the definition associated with the grade of two, and it seemed probable that another factor might have entered into the scoring. The markers might not have liked the sentimental nature of the topic that was discussed by the student. Sentiment often caused problems for markers because they seemed to approve or disapprove of the emotion that was expressed.

Marks for variable three, matters of choice, also ranged from two to five. One grader gave the score of two and it appeared that the marking guide was not followed carefully in this particular case. The composition did not match well with the definition for the grade of two.

In section two of the examination, a letter was written which discussed the need to renovate Nalwen's community hall. The letter was brief, being less than a page in length.

Marks for variable five ranged from one to four. The grader who awarded the score of four might have been influenced by the halo effect. The same grade had also been awarded for variables one, three, four, and for variable six through eleven by this marker.

In section three, question one, the student wrote that the proud girl in front of the class was trying to look like a flower but that her classmates were thinking of something else.

Marks for variable eight, thought and detail, ranged from two to five. Both extremes of the grades, the two and the five, seemed to be rather inconsistent. It appeared that descriptors could have been followed more carefully.

In section three, question two, the writer stated that the main idea was that "everyone thinks different from each other." The answer was two sentences in length.

Marks for variable ten, thought and detail, ranged from two to five. The grader who awarded the mark of five had awarded the same grade for variables eight and nine. The halo effect probably affected the scoring because the brevity of the answer and the generalizations expressed did not appear to match with the definition for the grade of five.

3. Generalizations Regarding Problem Papers

The problems that were analyzed in detail based upon each of the scoring variables in the reliability review papers were examined with a view to making generalizations about the difficulties that some markers had in awarding accurate grades. These problems clustered about a failure to apply the scoring descriptors.

A detailed study of the papers for which grading problems were evident revealed that failure to follow the scoring descriptors was the single underlying cause. Furthermore, this failure was the result of a halo effect; that is, there was a tendency among some raters to score all the variables high or low on the basis of a superficial opinion. When some markers liked or disliked something about a composition, they tended to reward or penalize the writer not only in the category relevant to the perceived merit or deficiency, but in other categories as well. This halo effect has several distinct aspects that deserve further description.

One of the most common examples of the halo effect resulted from a high number of mechanical errors. While the area of mechanics was intended to be included in the scoring variable labelled "matters of convention," for some readers the deficiency crept into all of the variables, resulting in low scores.

Another example involved raters' approval or disapproval of the subjects dealt with in the compositions. Views that the teachers liked, or with which they agreed, often resulted in high marks being awarded for all variables. Similarly, when papers expressed ideas that dealt with subjects teachers disliked, marks for all variables tended to be depressed.

Another example of the halo effect was produced when unusual compositions were written. For example, when students did not deal with the topic in a direct way but rather chose to write on a more symbolic level, some markers had trouble scoring the paper. Markers had difficulty with compositions that gave topic sentences nearer the end of the composition rather than at the beginning and with compositions that left material for the reader to infer. Some markers awarded higher grades when this occurred, but most markers who had problems penalized the piece of writing. This was especially noticeable when the grade of insufficient was given in that these markers had such problems with the inferential answer that they believed the composition did not deal with the assigned topic.

An additional example of the halo effect resulted when writers dealt with their topic in a sentimental fashion. Some markers rewarded and others penalized if a student wrote about themes that were emotional in nature. The greatest problems occurred for markers when the subject dealt with love and family relationships.

Another example involved the problem of different lengths of the compositions. Difficulties occurred both when compositions were brief and when they were long. Shorter compositions were both rewarded and penalized by markers; some individuals must have believed the short answer contained a concise, desired response while others must have felt that not enough was said about a subject. Longer answers were usually rewarded by giving a higher grade even when the response itself did not fit the assigned topic very effectively.

A less common example was related to false indicators of deep thought. Answers that contained complex wordings (even when the words were badly misused), sweeping statements, and over-generalizations were often rewarded by some of the markers.

A final example of the halo effect came from papers written by English as a second language students. When it was obvious that a student's paper had been written by an E.S.L. student, some markers penalized the composition in all scoring categories without utilizing the descriptor definitions effectively. Other markers who recognized an E.S.L. paper appeared to reward the writer by giving generous marks in categories other than the ones which dealt with mechanical errors. These markers seemed to adopt the attitude that E.S.L. students should be graded more leniently than other English 33 students because they were learning a second language.

These aspects of the halo effect seemed to be rather prevalent, and no doubt, were responsible for much of the inconsistency that existed among markers. A positive aspect of the problem was that the difficulties could be readily identified and, through training, it appeared likely that they could be reduced.

B. Reader-Based Problems

Problem readers were defined (see page 54) in terms of three criteria. The first criterion was a lack of agreement with other readers; the second was a tendency to grade too hard or too easy; and the third was a tendency to produce unusually high or low score spreads. In order to identify problem readers, the three criteria had to be operationally defined. Degree of agreement with other readers was taken to be the mean correlation between the grades awarded by each marker and every other individual marker. The severity of scoring was defined as the mean score given to all of the papers, and the score spread for each reader was taken to be the standard deviation of the scores submitted.

The relevant statistics, mean correlations, mean scores, and standard deviations, by reader, are outlined in Tables IV-2 and IV-3. These statistics serve as a basis for further study of reader-based problems according to the three criteria outlined above. The results are described under separate headings.

1. Lack of Agreement With Other Readers

A mean correlation of 0.8 between one reader and each of the others was set as the minimum value for satisfactory reliability of scoring. The value seemed to be a realistic expectation for two reasons. First, respected scholars had used this standard; and second, the situation present in this study was similar to that of other studies. For the purposes of this study, two levels of unsatisfactory reliability were defined. A mean correlation of less than 0.8 (values between 0.795 and 0.8 were all rounded to 0.8) but greater than or

Table IV-2

MEAN CORRELATIONS, MEANS, AND STANDARD DEVIATIONS FOR ALL READERS (JANUARY, 1986)

Reader #	Mean Correlation	Mean Score	Standard Deviation
1	.860	70.00	18.01
2	.815	64.14	17.75
3	.774	62.43	18.49
4	.869	67.33	23.89
5	.757	65.00	18.12
6	.662	66.50	12.72
7	.857	63.86	17.74
8	.721	61.86	13.37
9	.837	65.29	14.08
10	.775	70.14	21.83
11	.845	59.33	24.54
12	.722	69.71	11.49
13	.884	66.43	16.16
14	.794	63.71	15.06
15	.847	59.29	13.93
16	.815	68.29	14.31
17	.689	59.29	19.64
18	.773	71.29	10.09
19	.777	62.00	17.03
20	.864	60.14	16.34
21	.827	69.00	13.53
22	.808	61.29	14.97
23	.859	59.83	19.00
24	.862	68.86	18.01
25	.845	66.00	16.86
26	.880	65.17	18.58
27	.883	68.50	14.20
28	.724	60.67	18.55
29	.791	62.00	16.56
30	.901	69.67	16.97
31	.812	64.57	14.15
32	.878	60.57	17.80
33	.872	65.29	15.94
34	.877	69.67	14.31
35	.776	60.83	15.45
36	.797	58.29	17.88
37	.880	63.50	19.76
38	.737	64.71	19.88
39	.772	69.17	22.44
40	.836	63.43	15.69
41	.846	60.29	13.33
42	.462	53.29	21.71
43	.867	66.67	18.39
44	.836	60.33	26.85
45	.877	67.29	13.25

Table IV-2 (Cont'd)

Reader #	Mean Correlation	Mean Score	Standard Deviation
46	.824	67.86	17.66
47	.831	67.67	17.06
48	.853	66.29	17.39
49	.844	65.29	17.85
50	.825	71.29	16.62
51	.826	65.57	16.44
52	.744	70.43	17.98
53	.878	65.00	14.80
54	.792	74.00	13.35
55	.836	67.14	19.55
56	.838	67.67	12.88
57	.821	58.29	15.64
58	.826	66.14	18.33
59	.846	59.29	16.22
60	.793	66.57	20.57
61	.677	61.14	21.33
62	.867	66.00	21.14
63	.798	64.86	15.79
64	.848	65.67	16.63
65	.883	63.29	21.00
66	.821	65.57	16.65
67	.789	61.14	24.08
68	.866	60.17	15.64
69	.842	56.57	15.38
70	.647	56.57	16.41
71	.780	62.29	14.58
72	.884	63.00	16.00
73	.878	65.67	21.83
74	.863	68.14	16.85
75	.790	65.00	20.13

Overall Mean 64.51
Median 17.05

Table IV-3

MEAN CORRELATIONS, MEANS, AND STANDARD DEVIATIONS FOR ALL READERS (JUNE, 1986)

Reader #	Mean Correlation	Mean Score	Standard Deviation
1	.677	59.30	12.94
2	.777	68.60	13.95
3	.821	62.33	16.61
4	.825	65.80	16.57
5	.778	68.11	18.78
6	.699	67.90	15.39
7	.784	61.56	18.04
8	.841	63.67	16.96
9	.808	63.44	15.69
10	.808	65.33	14.81
11	.697	64.11	17.77
12	.739	63.67	19.27
13	.824	66.75	15.45
14	.802	62.56	14.37
15	.666	64.67	11.61
16	.723	64.25	13.27
17	.677	71.80	13.11
18	.698	55.22	15.94
19	.653	67.40	14.27
20	.521	70.38	12.45
21	.831	59.80	16.71
22	.784	66.00	23.54
23	.777	64.10	19.77
24	.816	63.80	12.62
25	.778	65.44	14.63
26	.799	64.67	16.37
27	.795	66.10	17.49
28	.769	69.11	13.15
29	.678	66.20	15.26
30	.457	63.90	16.04
31	.716	66.40	15.65
32	.796	65.90	12.89
33	.764	64.70	16.43
34	.805	68.90	19.87
35	.828	64.89	13.27
36	.739	63.60	18.57
37	.833	62.22	18.12
38	.753	58.30	12.45
39	.662	67.20	14.72
40	.697	65.90	11.25
41	.708	64.80	11.60
42	.771	66.80	13.76
43	.606	61.10	8.44
44	.814	63.44	16.09
45	.792	74.11	14.10

Table IV-3 (Cont'd)

Reader #	Mean Correlation	Mean Score	Standard Deviation
46	.793	68.00	18.74
47	.611	71.70	14.11
48	.809	62.30	13.94
49	.433	68.44	11.40
50	.666	62.88	11.78
51	.721	63.70	13.42
52	.752	60.50	10.29
53	.812	62.22	16.18
54	.615	65.11	14.89
55	.775	61.30	17.38
56	.769	63.10	16.85
57	.732	65.40	18.72
58	.782	66.89	15.05
59	.815	64.10	15.03
60	.853	65.75	20.52
61	.697	62.10	12.25
62	.642	63.20	13.18
63	.584	60.40	14.35
64	.778	61.50	15.83

Overall Mean 64.73
Median 15.06

equal to 0.7 was considered to be mildly unsatisfactory while values of less than 0.7 were considered to be definitely unsatisfactory. The markers in the middle level were judged to be in a gray area. It must be borne in mind that the mean correlations of all markers tended to be deflated by the generally low values contributed by the very unsatisfactory group. Highly satisfactory readers would not inflate the mean. Grades from good readers would correlate highly with grades from other good readers, but poorly with grades from poor readers. Poor readers, in turn, would not agree to any extent with anyone else. Also, because the sample of papers was small, it was possible that the mean correlations were low because of random error. It must be remembered, however, that in the reliability review situation readers were on their "best behavior" because they knew that they were accountable for the grades that they submitted. Mean correlations less than 0.7 represented readers who, for one reason or another, were far below expectation with respect to the reliability of scoring. Table IV-4 shows the results of the categorization of the readers.

Table IV-4

DISTRIBUTION OF CORRELATIONS BY CATEGORY

	Category			
Grading Session	Satisfactory r = 0.8 or >	Mildly Unsatisfactory r=0.7 to 0.795	Very Unsatisfactory r < 0.7	Total
January	51	19	5	75
June	19	25	20	64

2. Easy or Severe Markers

Mean scores of the markers over the reliability review samples were used to identify easy and severe readers. The degree of discrepancy to be tolerated was decided logically. A value of five was chosen; that is, rater means within five score points of the overall mean were judged to be acceptable, but discrepancies greater than five were not. Several factors influenced the choice. First of all, the letter grades B and C had score ranges of fifteen. A grade of B, for example, included score values from sixty-five to seventy-nine, and C ranged from fifty to sixty-four. A discrepancy of one third of a letter grade seemed reasonable, and certainly not too severe. A second consideration was directly related to the scores. A group of papers having a "true" score (the mean of all the scores) of sixty-five would receive grades ranging from sixty to seventy without objection from most teachers. Values that deviated further did not seem to be justifiable. A third consideration was that discrepancies exhibited by readers were probably lower than would occur in an actual scoring situation, because in the situation at hand, markers were motivated to perform at their best.

To assist in analyzing the extent to which deviant grading standards were a problem, different levels of discrepancy were established. Deviations of five to seven and one-half score points were judged "easy" or "hard." Deviations greater than seven and one-half points were judged "very easy" or "very hard." Table IV-5 shows the distribution of the deviations according to category.

Table IV-5

DISTRIBUTION OF DISCREPANCIES IN MEANS BY CATEGORY

		Category				
Grading Session	Overall Mean	V. Easy > 7.49	Easy 5 to 7.49	Satisf. 4.95 to -4.95	Hard -5 to -7.49	V. Hard <-7.49
January	64.51	1	8	57	6	3
June	64.73	1	3	57	2	1

3. Discrepant Spreads of Scores

With respect to standard deviations of scores submitted by readers on the reliability review sample, the range of discrepancy to be tolerated was arrived at through a logical approach based upon comparisons. First, frequency polygons representing the standard deviations for the January and June scoring sessions were plotted (Figure IV-1 and IV-2). Next, the acceptable range was defined as the values within five score points of the median value of the standard deviations. This range seemed reasonable in that, on the frequency polygons, the fringe values were cut off as unsatisfactory. While a discrepancy in the standard deviation of five score points might seem to be great, it should be borne in mind that the effect on grades of standard deviations of that magnitude is not serious.

A higher than usual standard deviation by a marker indicated that the spread of marks was greater, which meant

Figure IV–1
STANDARD DEVIATIONS FOR JAN.,1986 SCORING SESSION

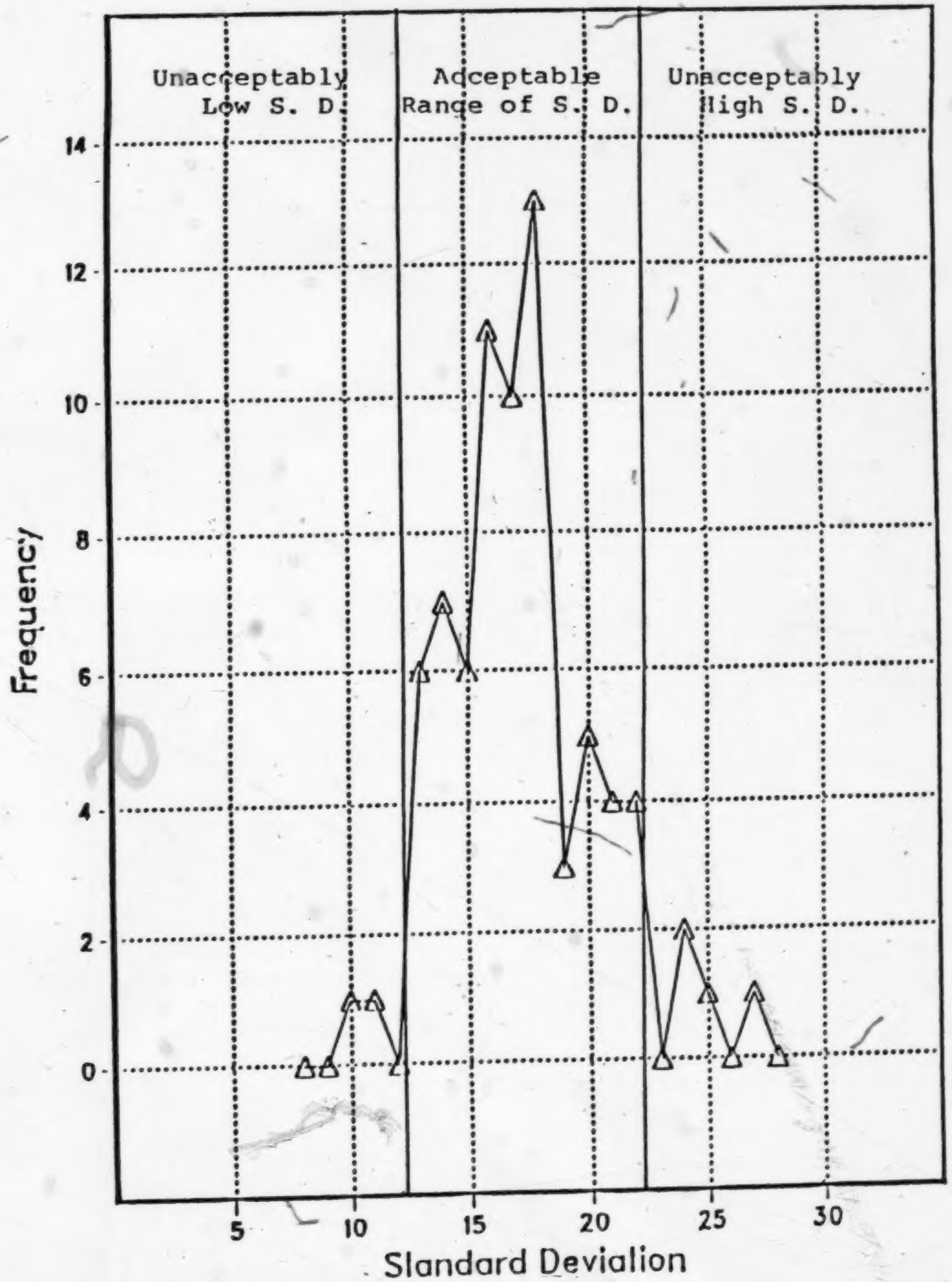
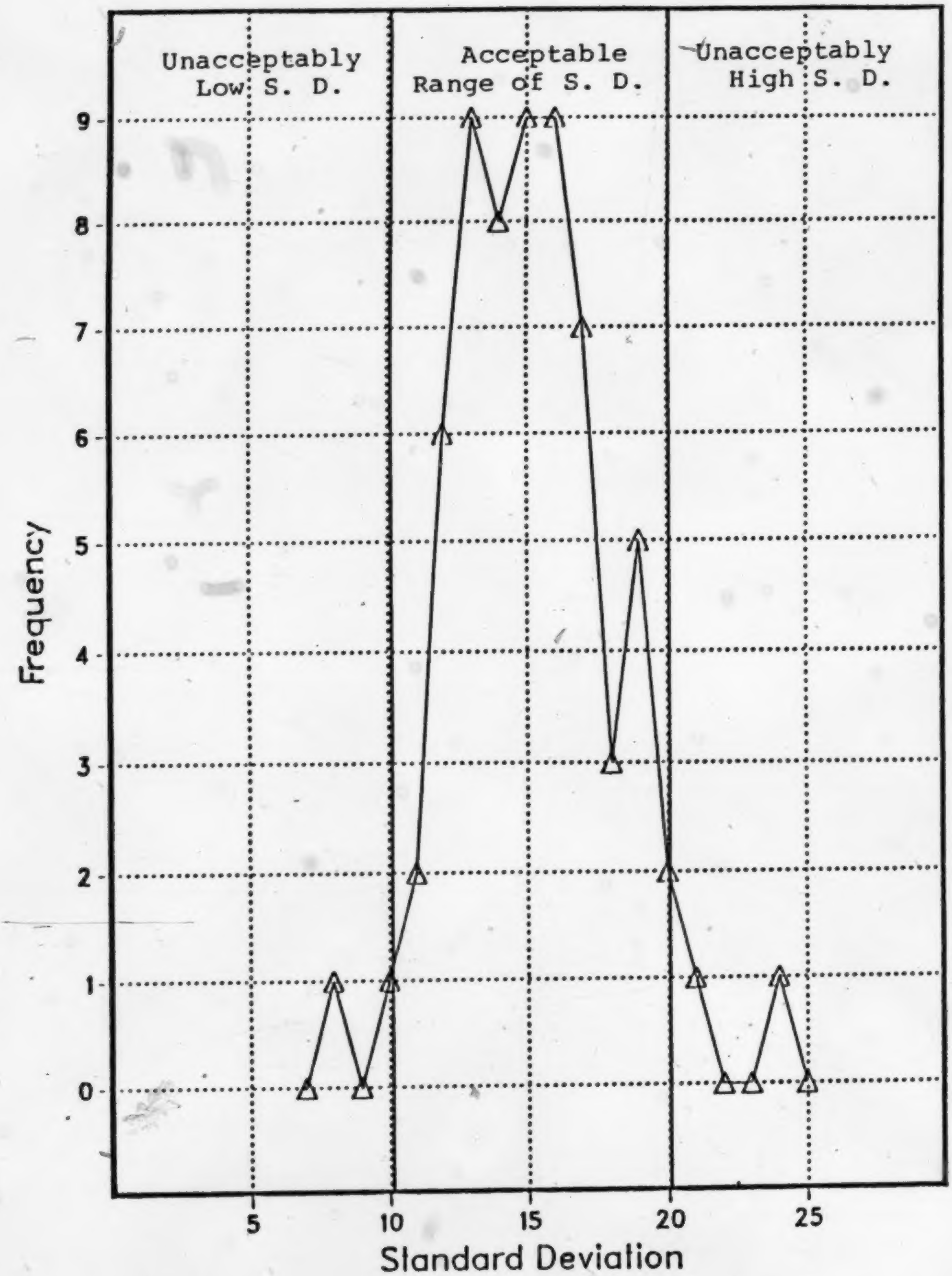


Figure IV–2
STANDARD DEVIATIONS FOR JUNE, 1986 SCORING SESSION



that good essays would receive higher marks than expected, and weak essays would receive marks lower than expected. The large number of scores in the middle range would be close to those awarded by non-discrepant readers.

4. Overall Results

A summary of the results of the study of the problem readers is given in Tables IV-6 and IV-7. Inspection of the figures in the tables revealed that the number of markers with unsatisfactory mean correlations was greater for the June, 1986, marking session than for the January, 1986, examinations. Forty-five of the sixty-four markers who participated in the reliability review scoring session in June had unsatisfactory correlations (below 0.8) while twenty-four of the seventy-five markers who worked in January had correlations below this standard. Twenty markers had correlations below 0.7 in June, of which four were below 0.6, while, for the January group, only five had correlations below 0.7, of which one was below 0.6.

With respect to the mean scores given to the seven reliability review papers from the January, and June, 1986, scoring sessions, the number of discrepancies (see Table IV-5) was not great; however, the patterns for the two sessions differed. The percentage of deviant markers was greater for the January session (24.0%) than for the June session (10.9%).

There were fewer problems with standard deviations than there were with the mean score and mean correlations. In

Table IV-6

PROBLEM READERS (JANUARY, 1986)

Reader #	Mean Corr.	Problem	Mean Score	Problem	S. D.	Problem
1	----	----	70.00	easy	----	----
3	.774	mild	----	----	----	----
4	----	----	----	----	23.89	high
5	.757	mild	----	----	----	----
6	.662	severe	----	----	----	----
8	.721	mild	----	----	----	----
10	.775	mild	70.14	easy	----	----
11	----	----	59.33	hard	24.54	high
12	.722	mild	69.71	easy	11.49	low
14	.794	mild	----	----	----	----
15	----	----	59.29	hard	----	----
17	.689	severe	59.29	hard	----	----
18	.773	mild	71.29	easy	10.09	low
19	.777	mild	----	----	----	----
28	.724	mild	----	----	----	----
29	.791	mild	----	----	----	----
30	----	----	69.67	easy	----	----
34	----	----	69.67	easy	----	----
35	.776	mild	----	----	----	----
36	----	----	58.29	hard	----	----
38	.737	mild	----	----	----	----
39	.772	mild	----	----	22.44	high
42	.472	severe	53.29	v. hard	----	----
44	----	----	----	----	26.85	high
50	----	----	71.29	easy	----	----
52	.744	mild	70.43	easy	----	----
54	.792	mild	74.00	v. easy	----	----
57	----	----	58.29	hard	----	----
59	----	----	59.29	hard	----	----
60	.793	mild	----	----	----	----
61	.677	severe	----	----	----	----
67	.789	mild	----	----	24.08	high
69	----	----	56.57	v. hard	----	----
70	.647	severe	56.57	v. hard	----	----
71	.780	mild	----	----	----	----
75	.790	mild	----	----	----	----

Table IV-7

PROBLEM READERS (JUNE, 1986)

Reader #	Mean Corr.	Problem	Mean Score	Problem	S. D.	Problem
1	.677	severe	59.30	hard	----	----
2	.777	mild	----	----	----	----
5	.778	mild	----	----	----	----
6	.699	severe	----	----	----	----
7	.784	mild	----	----	----	----
11	.697	severe	----	----	----	----
12	.739	mild	----	----	----	----
15	.666	severe	----	----	----	----
16	.723	mild	----	----	----	----
17	.677	severe	71.80	easy	----	----
18	.698	severe	55.22	v. hard	----	----
19	.653	severe	----	----	----	----
20	.521	severe	70.38	easy	----	----
22	.784	mild	----	----	23.54	high
23	.777	mild	----	----	----	----
25	.778	mild	----	----	----	----
27	.795	mild	----	----	----	----
28	.769	mild	----	----	----	----
29	.678	severe	----	----	----	----
30	.457	severe	----	----	----	----
31	.716	mild	----	----	----	----
33	.764	mild	----	----	----	----
36	.739	mild	----	----	----	----
38	.753	mild	58.30	hard	----	----
39	.662	severe	----	----	----	----
40	.697	severe	----	----	----	----
41	.708	mild	----	----	----	----
42	.771	mild	----	----	----	----
43	.606	severe	----	----	8.44	low
45	.792	mild	74.11	v. easy	----	----
46	.793	mild	----	----	----	----
47	.611	severe	71.70	easy	----	----
49	.433	severe	----	----	----	----
50	.666	severe	----	----	----	----
51	.721	mild	----	----	----	----
52	.752	mild	----	----	----	----
54	.615	severe	----	----	----	----
55	.775	mild	----	----	----	----
56	.769	mild	----	----	----	----
57	.732	mild	----	----	----	----
58	.782	mild	----	----	----	----
60	----	----	----	----	20.52	high
61	.697	severe	----	----	----	----
62	.642	severe	----	----	----	----
63	.584	severe	----	----	----	----
64	.778	mild	----	----	----	----

January, 1986, five markers had high standard deviations and in June, 1986, two markers had high standard deviations. Two markers in January and one marker in June had low standard deviations; these individuals were grading toward the mean.

It should be noted that mean correlations caused the greatest problem in the June scoring session and that there were fewer problems with respect to the means and standard deviations. In all but one case, marks with unsatisfactory means and standard deviations were also lacking with respect to the mean correlation. During the January scoring session, the results were more scattered.

C. Problems Related to the Scoring Variables

Problems based in the scoring variables were investigated through a sample made up of readers who had been involved in both the January, and June, 1986, scoring sessions. This group consisted of thirty-seven markers. Combining the two scoring sessions resulted in seventeen reliability review essays being rated by each marker. First, the correlation between each pair of readers, for each of eleven scoring variables was determined. Next, the mean correlation for each reader for each variable was calculated. Finally, the mean of all correlations for each of the variables was found. The magnitude of this correlation was taken as an indicator of the effectiveness of the variable. The results are shown in Table IV-8. Also shown is the range of mean correlations for each variable, and the rank order of these correlations.

Table IV-8

MEAN, RANGE, AND RANK ORDER OF CORRELATIONS BY SCORING VARIABLE

(JANUARY, AND JUNE, 1986)

Variable & Writing Task (#)	Mean Corr.	Range	Rank Order
1 (I) Thought & Detail	.501	.274 - .635	9
2 (I) Organization	.483	.209 - .596	10
3 (I) Matters of Choice	.690	.469 - .774	5
4 (I) Matters of Convention	.733	.587 - .817	1
5 (II) Thought & Detail	.537	.226 - .686	7
6 (II) Organization	.456	.037 - .596	11
7 (II) Writing Skills	.720	.489 - .829	3
8 (III,1) Thought & Detail	.512	.263 - .653	8
9 (III,1) Writing Skills	.694	.578 - .774	4
10 (III,2) Thought & Detail	.590	.295 - .730	6
11 (III,2) Writing Skills	.723	.597 - .815	2

A study of Table IV-8 produced some pertinent observations. First, some of the correlations ranged so low as to be not significantly different from zero. A correlation of 0.48 was required for significance at $p = .05$.

The mean correlation for scoring variable six (organization, functional writing task two) was not significant, and for variable two (organization, personal response to literature writing task one), the mean was barely significant. This meant, of course, that approximately half of the correlations for these two variables did not differ significantly from zero. It was noted, also, that for variables four (matters of convention, writing task one), seven (writing skills, writing task two), nine (writing skills, writing task three) and eleven (writing skills, writing task three), only, all of the correlations were significant.

It was disturbing but perhaps not surprising to note that the variables more closely related to the higher skills involved in writing--variable one, five, eight, and ten (thought and detail), and two and six (organization)--did not produce the strongest correlations. This distinction was left to variable four (matters of convention) and to variables seven and eleven (writing skills), which consisted of less sophisticated skills such as mechanics, diction, and sentence variety. It was a little encouraging to note that thought and detail, which was weighted heavily in the scoring, at least, did not produce the lowest correlations. These came from the variable labelled organization.

D. Problems Related to the Writing Tasks

Problems related to the writing tasks were investigated according to the proposed procedures (page 55). For each of the three writing tasks, the correlations between all pairs of readers were determined, then the mean correlation for each marker was calculated. These values are shown in Appendix D, Tables D-1 and D-2. Next, the mean of all these means was found for each of the tasks. This overall mean was taken as an estimate of the reliability of scoring for each writing task. The results are shown in Table IV-9.

Table IV-9

MEANS OF CORRELATIONS BETWEEN READERS BY WRITING TASK

	Task 1	Task 2	Task 3
January Means	.740	.672	.835
June Means	.689	.591	.695

Table IV-9 shows a consistent pattern for the two writing sessions. Correlations for task two were substantially lower than the values for tasks one and three in both January, and June, 1986. Task three had noticeably higher correlations than did the other tasks.

Figures IV-3 and IV-4 are frequency distributions of the correlations shown in Tables D-1 and D-2 (see Appendix D). Further consistencies in the patterns between the January and June writing sessions are revealed. All of the correlation distributions were skewed to the left; that is, each distribution was drawn out at the low end of the scale. This suggested that a few readers, perhaps ten percent of them, had problems with one or more of the writing tasks.

An obvious question, regarding Figure IV-3 and IV-4 is whether the same markers were consistently found among the

Figure IV-3
DISTRIBUTION OF MEAN CORRELATIONS FOR READERS OVER WRITING TASKS (JAN.,1986)

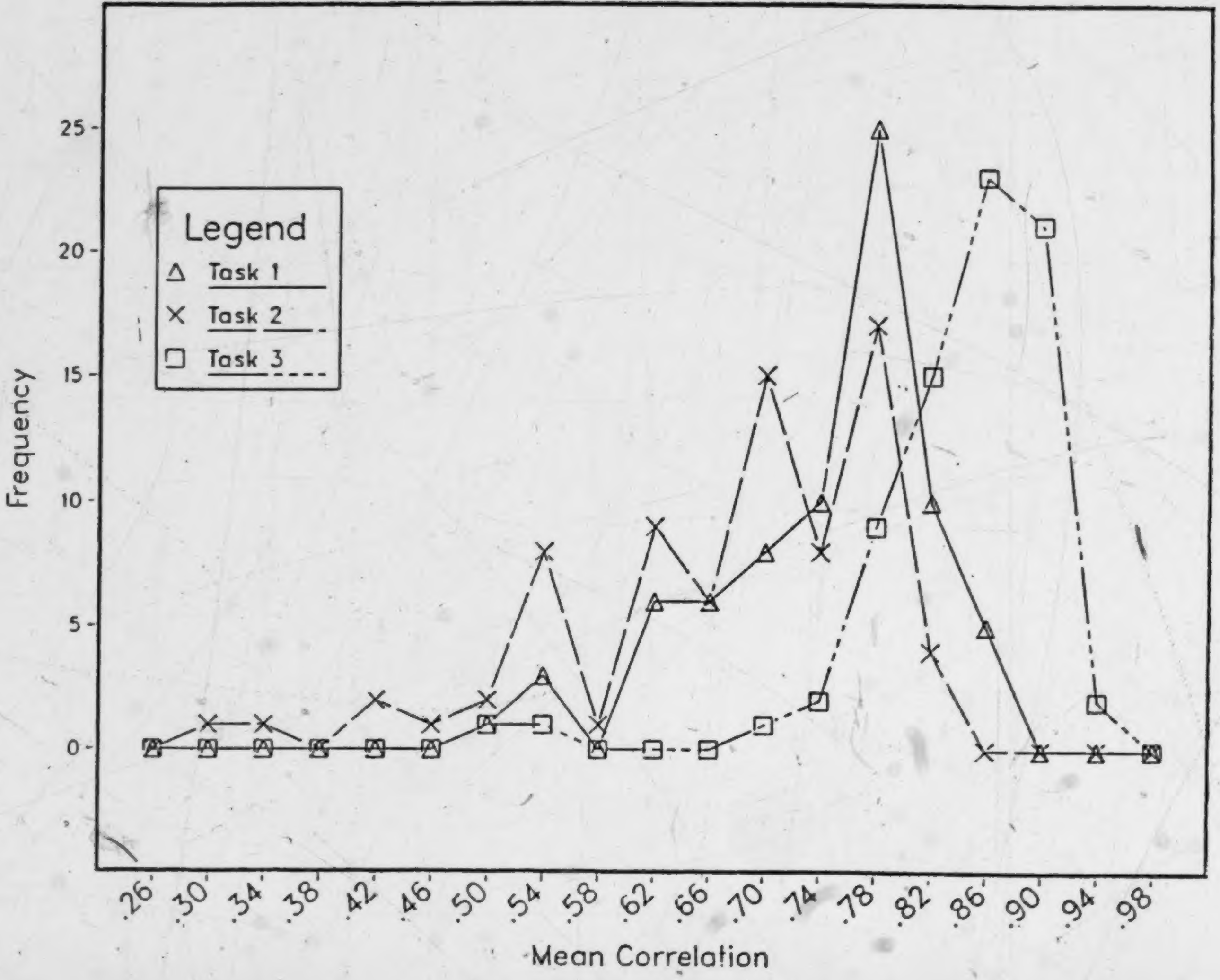
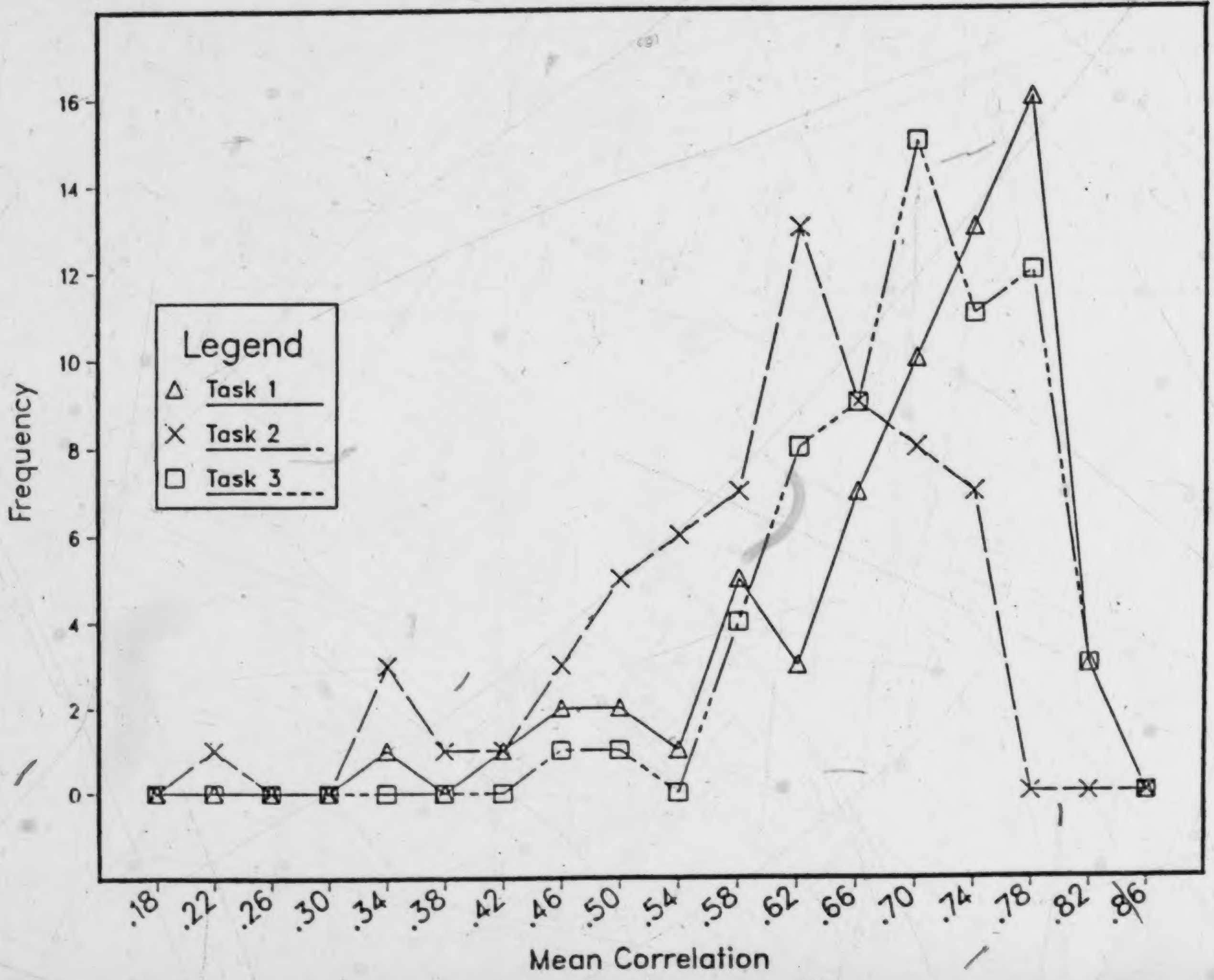


Figure IV–4
DISTRIBUTION OF MEAN CORRELATIONS FOR READERS OVER WRITING TASKS (JUNE, 1986)



weaker or the stronger readers for each of the three writing tasks. Tables D-1 and D-2 were used to locate the ten lowest and ten highest readers in each distribution. A comparison of the lists of markers revealed that there was, indeed, considerable overlap. A summary of the results is shown in Table IV-10.

Table IV-10

NUMBER OF TIMES READERS' MEAN CORRELATIONS APPEARED IN TOP OR BOTTOM TEN

Session	Category	Frequency of Occurrence 3	 2	 1
January	Lowest Ten	2	4	16
January	Highest Ten	2	5	14
June	Lowest Ten	2	3	18
June	Highest Ten	1	6	15

An explanation of the table by way of example is offered. For the January scoring session, two readers were found in the bottom ten on all three of the writing tasks, four readers were in the bottom ten on two of the tasks, and sixteen were in the lower ten on only one of the tasks.

II. Assessing Test Reliability

The reliability of the essay test as a whole, a broader measure than the reliability of scoring, was estimated through a measure of internal consistency, the alpha coefficient. During the January, 1986, marking session, a random sample of ninety-six papers was drawn after they had been routinely read by three markers. There were seventy-five markers involved; therefore, only a few of them scored more than one of the sample papers in each instance. It was possible, also, that a particular reader might not have scored any of them.

The alpha coefficient dealt with the eleven scoring variables employed by the readers. This method essentially processed the data as if each variable were an item of a test, then determined the internal consistency of the test. This is analogous to calculating the internal consistency of an exam made up of eleven written questions, and remotely resembles determining KR20 for an eleven-item multiple choice test.

The alpha coefficient is an estimate of test reliability, and should approximate the value that would be obtained if two similar essay examinations were administered to a group of students and the scores for the students correlated. It should be noted that disagreement among readers is one of the sources of unreliability, along with differences among reading tasks, differences among reader perceptions, etc.

The alpha coefficient for the first reading of the papers was calculated, and found to be 0.796. This meant that if the examinations had been read once by the seventy-five trained readers, the test reliability would have been 0.796. This value is essentially up to the minimum of 0.8 mentioned earlier (page 56). It is quite respectable when it is borne in mind that the 0.8 standard was for scoring reliability, only.

An important aspect of scoring of the sample of ninety-six papers was to determine the improvement that was achieved through scoring the papers three times, then calculating a final score. This was done by computing the alpha coefficient of the test by using the final scores awarded in the eleven variables after three scorings. A comparison was then made with the alpha value for the first scoring, only.

The alpha coefficient, based upon the final scores given for the eleven variables was 0.816, an improvement of 0.02. This modest improvement, and more significantly, the final value of 0.816 are important results in view of the other factors operating. These are discussed in Chapter V.

Discrepancies between first grading scores and final scores were calculated. Absolute differences were recorded; however, it should be noted that in fifty-eight instances, the final mark was higher than the initial score, and in only twenty-nine cases were the marks lowered. Table IV-11 represents a frequency distribution which showed absolute differences between the initial scores and final grades.

In response to the question of how large a discrepancy could be tolerated, the criterion of one third of a letter grade, applied previously (see page 119) was employed. A difference of more than five score points was judged undesirable. Application of this standard indicated that thirty-three of the scores (thirty-four percent) awarded in the first instance were unsatisfactory. Once again, the significance of this is left to the next chapter.

Table IV-11

ABSOLUTE DIFFERENCES BETWEEN INITIAL SCORES AND FINAL GRADES

Difference	Frequency	Difference		Frequency
22	1	10		1
21	0	9		5
20	0	8		6
19	0	7	Unacceptable	5
18	1	6	Difference	6
17	1			
16	0	5	Acceptable	8
15	2	4	Difference	11
14	0	3		8
13	2	2		13
12	0	1		14
11	3	0		9

Summary

Data based upon the reliability review samples were analyzed to identify, first, problems related to the compositions. It was found that failure on the part of some readers to follow the scoring descriptors was the single cause of reduced reliability. This cause identified as a halo effect had several aspects: a large number of mechanical errors, approval or disapproval of the views expressed by the writer, unusual compositions, sentimental or emotional content, length of the essay, false indicators of deep thought, and evidence that English was the writer's second language.

The second problem studied related to the readers. Three aspects were investigated: readers whose grades had low correlations with scores of others, those who consistently gave high or low grades, and those whose scores had wider spreads than others. For the January scoring session, twenty-four of the seventy-five readers were judged to be at least mildly unsatisfactory, while for the June session forty-five of the sixty-four were so judged. With respect to consistently high or low grades, eighteen of the seventy-five readers were unsatisfactory in January, and seven of the sixty-four were deficient in June. Unacceptable score spreads occurred in seven of the seventy-five readers in January, and in three of the sixty-four markers in June. It was noted, finally, that there were more problems associated with readers in the June session as compared to the January session.

The third problem studied related to the eleven scoring variables. Data were based upon the readers (thirty-seven) who were involved in both the January and June scoring sessions. The pattern that emerged was that the variables involving simpler skills, such as matters of convention, were scored most reliably, while those that dealt with more complex matters, such as organization, were weakest.

Finally, problems related to the writing tasks were studied. It was quite evident that, for both scoring sessions, the third composition was scored most reliably, and the second least reliably. It was also noted that the distribution of correlations in all cases was skewed to the left. This suggested that for each writing task, a few of the readers gave scores that did not agree with the consensus.

The reliability of the essay as a whole (alpha coefficient) was calculated based upon the first reading and then upon the final marks that were awarded. The values were 0.796 and 0.816 respectively. Discrepancies between the first reading scores and final grades were also calculated. For thirty-three of the ninety-six readers, the discrepancies were unacceptably large.

CHAPTER V

CONCLUSIONS, RECOMMENDATIONS, AND IMPLICATIONS

The results listed in Chapter IV suggest conclusions, recommendations, and implications related to the reliability of scoring and to the improvement of test reliability in general. These are dealt with in this chapter, first, under the headings used previously, that is, problems related to the compositions, readers, scoring variables, and writing tasks. In addition, the overall test reliability is examined in an attempt to assess the effectiveness of the scoring procedures that were used.

I. Compositions Having Low Scorer Reliability

A. Problems Related to the Compositions

Problems based in the compositions caused some markers to produce discrepant or inaccurate grades. The composition-based problems mentioned in Chapter IV were aspects of a single cause, a halo effect, that resulted in readers failing to follow the scoring descriptors. The problems were: numerous mechanical errors, readers disagreeing with views expressed by the writers, unusual compositions, a sentimental or emotional approach, unduly long or short compositions, false indicators of deep thought, and expression typical of English as a second language students.

The implications for improving scoring seem rather clear. Training of the markers consisted primarily of reviewing in detail the scoring descriptors and of applying the descriptors to selected papers used for training. The reliability review papers employed during the scoring sessions were used as a check on the application of the scoring descriptors. The instructional procedures, however, did not deal with special problems that occurred such as those described here because there was little or no awareness of them. It follows, then, that scoring reliability should improve if, during the instructional sessions, the readers are made particularly aware of the problems revealed in this study.

A basic assumption of this study was that writing could be separated into different compartments called variables for assessment purposes. One explanation of the difficulty some readers had in accomplishing this intent is that these readers allowed the halo effect to influence their judgment. A second possibility which was not examined in this study is that the very nature of writing is such that factors such as thought and detail, organization, and matters of choice are so interrelated that they cannot be isolated.

> **Recommendation One:** In the marker training sessions, the readers should be made aware of the composition-based problems that lead to the halo effect and the failure to follow scoring descriptors. These problems include either unduly positive or negative reaction to: numerous errors in mechanics, views expressed by the writers, unusual compositions, sentimental or emotional

content, brief or lengthy essays, false indicators of deep thinking, and expression typical of English as a second language students.

B. Reader-Based Problems

It was noted earlier that a relatively large proportion of the markers in both the January, and June, 1986, scoring sessions were considered deficient with respect to agreement with other readers. The problem regarding hard or lenient markers was less severe, as was the problem of markers who had discrepant score spreads. The fact that the lack of agreement was a more serious problem in June as compared to January was not readily explainable. Several possible contributors, however, came to light when the matter was discussed with Alberta Education supervisors.

First of all, there was more pressure on the markers in June. There were more papers, fewer markers, and the marking period stretched over a longer period of time. Consequently, the fatigue factor was greater in June than in January. An additional contributor to fatigue involved the timing of the scoring. The June scoring session commenced five days after the end of the school year. As a result, the energy of the teacher-markers was likely at a lower level in June than in January.

Another contributing factor to the lower level of agreement among markers in June might have been caused by a last-minute shift of markers that became necessary. When the marking session was about to begin, it was found that additional markers were needed for English 30, the grade twelve course intended for matriculation students. A call went out to those English 33 markers who were also qualified to score English 30 to leave the English 33 group. As a result, ten people were shifted, which further increased the work load for the English 33 scoring groups.

The fact that a considerable proportion of the readers were deficient with respect to agreement with others indicated that special attention should be given to this problem. The practice of scoring reliability review papers could be extended to serve this purpose. A closer study of the results, with more emphasis placed upon the necessity of achieving higher agreement with other markers, could be conducted without serious difficulty. Also, a policy could be developed whereby the overall reliability of each marker was recorded. In this way, graders with favorable statistics could be urged to serve as readers more frequently than others. In addition, a few readers who achieved very weak statistics might not be engaged as readers in future scoring sessions.

Monitoring of tendencies to grade hard or easy or to give unduly high score spreads should also be conducted. If additional instruction to reduce these problems were unsuccessful, it might be necessary to resort to mathematical procedures for equating reader standards. The procedures are simple. Scores by hard or easy markers are adjusted by adding or subtracting the difference between their mean and the overall mean on a sample of papers. For example, if a sample of papers is drawn (twenty-five papers), and these are scored independently by all the markers, a mean score for every reader could be determined, as could an overall mean. If one of the raters had a mean that was 2.5 below the overall mean, then 2.5 would be added to every score awarded by that reader.

A marker whose standard deviation was half of the means of all the standard deviations resulting from scoring the sample would be corrected by first doubling every one of his or her scores, then subtracting so as to correct for the mean. When all the scores are doubled, the standard deviation is also doubled. The mean, however, is also doubled and is therefore too high. Subtracting a fixed amount from each score, as described in the previous paragraph, would correct this fault.

In view of the findings that some markers were highly reliable raters while some were not, and that some graders gave scores that were too high, too low, or had unduly high or low standard deviations, two recommendations were formulated.

> **Recommendation Two:** A core group of markers who have a history of grading reliably should be identified. These people could be given special recognition and encouraged to serve as scorers on frequent occasions. Also, markers who grade unreliably should be given special instruction.

Readers who fail to come up to standard should not be engaged again.

Recommendation Three: Mathematical adjustment of scores should be used to correct tendencies to grade hard or easy or to give scores with unusually high or low standard deviations when the situation warrants.

C. Problems Related to the Scoring Variables

It was noted earlier (page 128) that the lower correlations were associated with the more complex writing skills such as organization and thought and detail. The higher correlations were related to the simpler skills, such as matters of choice and matters of convention. For these variables, expected levels of performance were more easily defined.

The variables yielding the lower scoring reliabilities are essential to the judging of a written product. Thought and detail, for example, cannot be disregarded, downgraded in importance, nor modified to any great extent without damaging the validity of scoring, that is, the extent to which the variables together really represent skills in writing. Other means for improving the performance of the variables must be found.

Identifying the troublesome variables is a first step toward improving the scoring reliability. These variables could be given special attention when readers were being trained, and when reliability checks were being conducted.

Reliability of scoring of each variable could also be checked. It is possible that a few objective standards will evolve for the variables that present difficulties with respect to consistency of scoring.

> **Recommendation Four:** Special attention should be given to instructing and monitoring the scoring of each of the variables.
>
> **Recommendation Five:** More objective standards for scoring the higher process variables should be sought. The emphasis should be upon getting a consensus through discussion and through following the scoring descriptors, then insisting that grading be done according to the standards set up by the group, not by the individual rater.

D. Problems Related to the Writing Tasks

The results of the investigation of the problems related to the writing tasks, conducted by comparing mean correlations across markers for each of the three tasks, revealed that writing task two (functional writing) was scored less reliably than the other two assignments. It was further noted that the reliability for task three was highest. Differences in reliability among the three tasks could be attributed to the combination of variables present and/or differences in the nature of the writing task.

It was immediately evident that task two suffered with respect to reliability because of the presence of the variable labelled organization. In Table IV-8 (page 127), it was reported that variable six, organization, had a mean correlation between pairs of readers of 0.456, the lowest of all. This same variable, in reading task one (variable two), also had a weak correlation (0.483). The variable was not included in the grading of writing task three. This fact no doubt contributed to writing task three being scored the most reliably of the three. Writing task one included the problem variable, but it made up a smaller proportion of the total score than was the case for task two. It seemed clear, then, that the variable of organization was a strong factor in the lower scoring reliability of task two.

The hypothesis that some trait of writing task two contributed to reduced scoring reliability could not be substantiated from these data. The other scoring variables involved (thought and detail and writing skills) were scored as reliably in task two as they were in the other writing assignments. It must be concluded, therefore, that the presence of the variable labelled organization was the primary and perhaps the sole cause of the poorer performance in scoring.

The solution to this problem of low reliability of scoring of students' ability to organize might be simple. If recommendation five were acted upon and the procedures successful, the difficulty found in scoring the functional writing task (task two) should be alleviated; however, a reconsideration of the variables assessed in each of the writing tasks should not be ruled out. It seemed strange, at first glance, that the variable labelled organization was included in the grading of writing task one and two, but not in task three. There was, however, a good reason for its omission. The responses elicited in task three tended to be short, and therefore required a minimum of organization. Eliminating the category called organization from task two might be considered; however, this procedure smacks of sacrificing in the area of validity for the sake of improving reliability, and this is not acceptable. If this procedure were carried to its ultimate conclusion, all variables would be eliminated save the one for which reliability of scoring was highest. This would no doubt result in high scorer reliability, but only a small and rather insignificant aspect of essay writing would be evaluated. A better procedure might be to increase the writing required for task three and then include organization as one of the scoring variables. This change would increase the importance of this variable and would make the writing tasks more comparable, but would reduce scorer reliability; however, if special attention were given to scoring this variable, the problem should be alleviated.

> **Recommendation Six:** The variables to be included for scoring in each of the writing tasks should be reconsidered. In particular, the variable labelled organization should be included in all three of the writing tasks, and special attention should be given to improve its reliability of scoring.

II. Assessing Test Reliability

The results of the calculation of overall test reliability were rather encouraging. Table II-1 (pages 37 and 38) revealed that in approximately one third of the studies reviewed, a scoring reliability of 0.8 or better was achieved. The reliability of approximately 0.8 recorded in this study was especially satisfying when it was borne in mind that other sources of unreliability besides scorer disagreement were present. It was realized, of course, that reliability coefficients computed by different methods were comparable in a rough way, only. Direct comparisons with other studies were not possible because no other studies employing the alpha coefficient were found.

Recommendations made so far in this study aimed at improving scorer reliability would, if successful, also improve the test reliability. It seems safe to say, then, that the procedures employed by Alberta Education produced respectable results insofar as reliability was concerned.

The matter of the improvement in test reliability brought about by scoring the papers more than once deserves further comment. The procedure resulted in a small increase in test reliability. This was encouraging in view of the fact that the halo effect would tend to inflate the alpha coefficient computed for marks based upon the first reading, but not for the final marks. This may be inferred from the fact that the alpha coefficient is a function of the correlations between the variables. A halo effect would increase these correlations and therefore elevate the alpha value. For final marks, no halo effect is present because the score for each variable is based upon a consensus by all of the markers. While the alpha coefficients might be judged to be satisfactory, the question of whether the increase that resulted from repeated scoring was worth the expenditure of money, time, and effort remained.

Before reaching a decision on whether the repeated scoring of the tests is worthwhile, an additional factor must be considered, that is, the effect upon individual students. In the sample, there were several instances where the final grades differed substantially from the marks awarded in the first reading. The differences between initial scores and final scores, revealed in Table IV-11 (p. 135) indicate that about one-third of the marks awarded initially were at least mildly unsatisfactory, and a few were highly erratic. If deviant marks awarded in the first instance were not adjusted as a result of subsequent readings, the students would suffer. Inordinately low marks on the first reading, if uncorrected could, obviously, affect the students negatively. It is less obvious that undeserved high marks that go uncorrected could also harm the students. The pupils themselves, however, would have inflated ideas of their achievement as would those who were judging the potential for future success. In any case, then, grades that did not reflect reality were potentially harmful. In order to help ensure the accuracy of grades for individual students, it would appear that repeated scoring of the essay papers is not merely desirable, but essential to the scoring process.

Recommendation 7: The procedure of scoring papers more than once should be continued.

The calculation of an alpha coefficient based upon a sample of papers drawn after the scoring is completed is a relatively simple and inexpensive process. There is no reason why it should not be routinely carried out after every scoring session so that comparisons can be made over time. The results would provide an ongoing check on the reliability of the English 33 test.

Recommendation 8: The alpha coefficient, based upon a random sample, should be computed at the conclusion of each scoring session.

III. Implications

This study is of value and interest to any organization involved in grading compositions, on a large-scale, at the school-leaving level. Results apply especially to Alberta Education for the obvious reason that their data and their system of scoring served as the base for the research. However, the identification of the halo effect has implications for all grading of high school compositions. The suggested methods of controlling this problem are of immediate value to the Alberta examination system, but are generalizable to other systems to the extent that the assigned tasks and the scoring procedures are the same. Retention of a core of proven readers, monitoring the reliability of scoring throughout the marking session, and mathematical adjustment of readers' scores can be used to advantage in Alberta and probably in other jurisdictions.

Suggestions for sharpening the variables through instruction and consensus in definitions after discussion is of particular value to Alberta, as is the recommendation that the writing tasks employ more uniform variables. These suggestions are also of value to all large-scale scoring of written work if the procedures used resemble those employed by Alberta Education.

The importance of having the papers graded several times is underscored, and judged to be necessary in the Alberta situation. This method has been advocated by other researchers (Coffman, 1971b, Akeju, 1972, Diederich, 1974), but this study provides additional hard data to back up the recommendation. This procedure serves to reduce or eliminate the halo effect, and to a large extent, correct the errors in judgment that individual markers make.

Over the long term, it may be that the use of the alpha coefficient to estimate the reliability of an essay examination as a whole will become widespread. No instances were found in the literature where this important matter was addressed; rather, previous studies focused upon the more limited topic of the reliability of scoring. The alpha coefficient can be applied in a situation where essays are graded on many variables, and where a sizable sample of papers can be drawn. The conclusion that an inflated value of the coefficient caused by having only one rater grade each paper is also a valuable contribution. For the Alberta scene, the research involving the alpha coefficient is of immediate value. It provides details of a procedure that can be applied at once, and also reveals that the reliability of the Alberta English 33 examination is quite respectable.

IV. Recommendations for Further Research

This study on the reliability of essay grading, by nature, was limited to a specific situation in that compositions written by English 33 students under examination conditions in Alberta were scored. The raters were Alberta-certificated teachers who had taught the course in the previous year. Important research possibilities outside of these limitations are suggested. Some of these are listed below.

1. Research on the reliability of scoring compositions written by English 30 (academic-stream) students should be conducted, and the results compared with those for English 33.

2. Reliability of scoring creative writing should be investigated. At present, skill in creative writing is one of the aspects considered by teachers of English when they compile marks representing students' achievement over the instructional term. The question arises as to how reliably teachers rate creative writing.

3. Research on why teachers score unreliably in some situations should be carried out. The question of whether

the halo effect noted in this study is related to personality traits of the raters needs to be answered.

4. The question of sex bias or differences arises. Research should be conducted to determine if there are effects attributable to the sex of the writers and the sex of the raters.

5. Longitudinal studies should be set up to monitor reliability of scoring over time. As examination procedures are adjusted over the years (for the length of examinations, number, and nature of topics, scoring variables, qualifications of readers, etc.) effects on reliability of scoring should be determined.

6. Further research on the reliability of examinations as a whole should be carried out. Methods other than the alpha coefficient should be used, and comparisons made.

7. Studies on the validity of English 33 and English 30 should be organized. While this study, like many others, focused upon reliability, the more comprehensive problem of validity should be addressed. Reliability and validity are not independent of one another, and high reliability is essential to high validity. Good reliability, however, does not ensure a high level of validity. A study to determine what can be predicted based upon English 33 and English 30 marks would be of real value.

The above list is not intended to be exhaustive, but is intended to give at least some direction for possible future research.

Summary

With respect to the reliability problems related to scoring, the causes of this halo effect, identified in Chapter IV, were reviewed and discussed. An argument was submitted that, since the causes of the problem had been revealed, instructions for the markers could focus upon them and reduce their effects. A recommendation in this vein was made.

Regarding reader-based problems, some readers were consistently high in reliability of scoring, while some were low. A recommendation was made that a core of reliable markers should be identified and encouraged to serve, and that weak markers should be given special instruction or not re-engaged. Mathematical adjustment of scores was recommended to correct for easy or severe grading, or scoring by readers that resulted in unduly high or low distributions.

The primary problem related to the scoring variables was identified as the tendency for the more complex writing tasks to be scored less reliably than the simpler tasks. For example, organization was scored less reliably than matters of convention. Recommendations for alleviating the problem included improved instruction and monitoring related to each variable, and more objective standards, arrived at through discussion.

The writing tasks varied in reliability of scoring because of differences in length and differences in the variables used for scoring. A recommendation that the variables should be reviewed and that the writing tasks should be more uniform resulted.

Calculation of alpha coefficients indicated, first, that the reliability of the Alberta written examination for English 33 was quite satisfactory, and second, that there was merit in having each paper scored more than once.

Implications of the study were stated. These focused upon the importance of the findings for Alberta, in particular, and for any situation involving large-scale scoring of high school written examinations.

Seven recommendations for further research were suggested.

BIBLIOGRAPHY

Akeju, S. A. "The Reliability of General Certificate of Education English Composition Papers in West Africa." Journal of Educational Measurement 9 (Summer 1972): 175-180.

Alberta Education. Grade Twelve Diploma Examination, English 33, Part A: Written Response, January, 1986. Edmonton, Alberta: Alberta Education, 1986.

Alberta Education. Grade Twelve Diploma Examination, English 33, Part A: Written Response, June, 1986. Edmonton, Alberta: Alberta Education, 1986b.

Anderson, Harold A. and Arthur E. Traxler. "The Reliability of the Reading of an English Essay Test--A Second Study." The School Review 48 (September 1940): 521-530.

Braddock, Richard. "English Composition," in Encyclopedia of Educational Research. 4th ed. Ed. Robert L. Ebel. Toronto: Collier-Macmillan Ltd., 1969.

Braddock, Richard, Richard Lloyd Jones, and Lowell Schoer. Research in Written Composition. Champaign, Illinois: National Council of Teachers of English, 1963.

Breland, Hunter M. The Direct Assessment of Writing Skill: A Measurement Review. New York: College Board Report No. 83-6, 1983.

Britton, J. N., N. C. Martin, and H. Rosen. Multiple Marking of English Compositions. London: Her Majesty's Stationery Office, 1966.

Caldwell, Otis W., and Stuart A. Courtis. Then and Now in Education, 1845-1923. New York: World Book Co., 1924.

Cast, B. M. D. "The Efficiency of Different Methods of Marking English Composition." British Journal of Educational Psychology 9 (1939): 257-269 and 10 (1940): 49-60.

Coffman, William E. Educational Measurement. Ed. R. L. Thorndike. 2nd ed. Washington, D. C.: American Council on Education, 1971a.

Coffman, William E. "On the Reliability of Essay Examinations in English." Research in the Teaching of English 5 (Spring 1971b): 24-36.

Cooper, Charles. "Measuring Growth in Writing." English Journal 64 (March 1975): 111-120.

Cooper, Charles R. The Nature and Measurement of Competency in English. Urbana, Illinois: National Council of Teachers of English, 1981.

Cooper, Charles R. and Lee Odell. Evaluating Writing: Describing, Measuring, Judging. Urbana, Illinois: National Council of Teachers of English, 1977.

Cooper, Charles R. and Lee Odell. Research on Composing--Points of Departure. Urbana, Illinois: National Council of Teachers of English, 1978.

Coward, Ann F. "A Comparison of Two Methods of Grading English Compositions." Journal of Educational Research 46 (October 1952): 81-93.

Darsie, Marvin L. "The Reliability of Judgments Based on the Willing Composition Scale." Journal of Educational Research 5 (January 1922): 89-90.

Diederich, Paul B. Measuring Growth in English. Urbana, Illinois: National Council of Teachers of English, 1974.

Diederich, Paul B., J. W. French, and S. T. Carlton. Factors in Judgments of Writing Ability (Research Bulletin RB61-15). Princeton: Educational Testing Service, 1961.

Drake, C. A. "An Examination-Marking Experiment." School and Society 52 (1940): 95-96.

DuBois, Philip H. A History of Psychological Testing. Boston: Allyn and Bacon, Inc., 1970.

Dunn, Thomas William Patrick. "Teachers' Perceptions of Writing Quality and Criterion Levels of Composition." Diss. U of Alberta, 1982.

Edgeworth, F. Y. "The Element of Chance in Competitive Examinations." Journal of the Royal Statistical Society 53 (1890): 460-475 and 644-673.

Eells, Walter C. "Reliability of Repeated Grading of Essay-Type Examinations." Journal of Educational Psychology 21 (1930): 48-52.

Englehart, Max D. "Examinations," in Encyclopedia of Educational Research. Ed. W. S. Munroe. New York: Macmillan Co., 1950.

Finlayson, Douglas S. "The Reliability of the Marking of Essays." British Journal of Educational Psychology 21 (1951): 126-134.

Follman, John C., and James A. Anderson. "An Investigation of the Reliability of Five Procedures for Grading English Themes." Research in the Teaching of English 1 (Fall 1967): 190-200.

Foster, David. A Primer for Writing Teachers. Upper Montclair, New Jersey: Boynton/Cook Publishers, Inc., 1983.

Gee, Thomas Walter. "Drafting and Revising Processes in Twelve Students' Examination Writing." Diss. U of Alberta, 1984.

Gee, Thomas Walter. "Drafting and Revising Processes in Twelve Students' Examination Writing." The Alberta Journal of Educational Research 2 (June 1987): 96-114.

Godshalk, Fred I., Frances Swineford, and William E. Coffman. The Measurement of Writing Ability. New York: College Entrance Examinations Board, 1966.

Guba, Egon, and Yvonna S. Lincoln. Effective Evaluation. San Francisco: Jossey-Bass Publishers, 1982.

Gulliksen, H. Theory of Mental Tests. New York: John Wiley, 1950.

Hartog, P., and E. C. Rhodes. An Examination of Examinations. London: Macmillan and Co., 1936.

Haskins, Charles Homer. The Rise of Universities. Ithaca: Cornell University Press, 1957.

Hillegas, M. B. A Scale for the Measurement of Quality in English for Young People. New York: Bureau of Publications, Teachers College, Columbia University, 1912.

Horner, Judith. "How Do We Know How Well Kids Write?" English Journal 67 (October 1978): 60-61.

Huddleston, Edith M. "Measurement of Writing Ability of the College Entrance Level: Objective vs. Subjective Testing Techniques." Journal of Experimental Education 22 (1954): 165-213.

Hudelson, Earl. Hudelson English Composition Scale. New York: World Book Company, 1923.

Hulten, C. E. "The Personal Element in Teachers' Marks." Journal of Educational Research 12 (June 1925): 49-55.

Johnston, Brian. Assessing English--Helping Students To Reflect on Their Work. Rosebery, New South Wales: Bridge Printery, 1983.

Judine, Sister M. A Guide For Evaluating Student Composition. Urbana, Illinois: National Council of Teachers of English, 1965.

Lees, Elaine O. Critical Monism, Critical Pluralism, and the Ideal of Inter-Rater Reliability. Pittsburgh: ERIC, 1981.

Marsh, Herbert W., and Robert Ireland. Multidimensional Evaluations of Writing Effectiveness. Sydney, Australia: ERIC, 1984.

McColly, William. "What Does Educational Research Say About the Judging of Writing Ability?" The Journal of Educational Research 64 (December 1970): 148-156.

Meredith, Vana Hutto, and Paul L. Williams. "Problem Identification and Control." Educational Measurement Issues and Practice 3 (Spring 1984): 11-15 and 35.

Myers, Albert E., Carolyn B. McConville, and William E. Coffman. "Simplex Structure in the Grading of Essay Tests." Educational and Psychological Measurement 26 (1966): 41-54.

Myers, Miles. A Procedure for Writing Assessment and Holistic Scoring. Urbana, Illinois: National Council of Teachers of English, 1980.

Myers, Miles. The Teacher-Researcher: How to Study Writing in the Classroom. Urbana, Illinois: National Council of Teachers of English, 1985.

Myers, Miles, and James Gray. Theory and Practice in the Teaching of Composition. Urbana, Illinois: National Council of Teachers of English, 1983.

Najimy, Norman C. Measure for Measure--A Guidebook for Evaluating Students' Expository Writing. Urbana, Illinois: National Council of Teachers of English, 1981.

Nyberg, Adell M. "A Study of the Reliability of an Alberta Composition Scale." Unpublished M.Ed. colloquium. U of Alberta, 1977.

Nyberg, Verner R., and Adell M. Nyberg. Alberta Essay Scales: Models. Edmonton, Alberta: Faculty of Education, Univeristy of Alberta, 1982.

Nyberg, Verner R., and Adell M. Nyberg. *Technical Report On Alberta Essay Scales: Models*. Edmonton, Alberta: Faculty of Education, University of Alberta, 1982b.

Odell, L., and C. R. Cooper. "Procedures for Evaluating Writing: Assumptions and Needed Research." *College English* 42 (September 1980): 34-43.

Popham, W. James. *Educational Evaluation*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

Russell, David H., Edmund J. Farrell, and Margaret J. Early. *Research Design and the Teaching of English*. Champaign, Illinois: National Council of Teachers of English, 1964.

Selltiz, Claire, Marie Jahoda, Morton Deutsch, and Stuart W. Cook. *Research Methods in Social Relations*. Toronto: Holt, Rinehart and Winston, 1951.

Scannell Dale P, and Jon C. Marshall. "The Effect of Selected Composition Errors on Grades Assigned to Essay Examinations." *American Educational Research Journal* 3 (March 1966): 125-130.

Smith, Vernon H. "Measuring Judgment in the Evaluation of Written Composition." *Research in the Teaching of English* 3 (Fall 1969): 181-195.

Stalnaker, John M. "The Construction and Results of a Twelve-Hour Test in English Composition." *School and Society* 39 (February 17, 1934): 218-224.

Stalnaker, John M. "Essay Examinations Reliably Read." *School and Society* 46 (November 20, 1937): 671-672.

Starch, Daniel, and Edward C. Elliott. "Reliability of the Grading of High School Work in English." *School Review* (September 1912): 442-457.

Student Evaluation. *Grade 12 Diploma Examinations Program English 30 and English 33 1985-86 School Year*. Edmonton, Alberta: Alberta Education, 1985.

Student Evaluation. *Provincial Report Grade 12 Diploma Examinations January 1986 Administration*. Edmonton, Alberta: Alberta Education, 1986a.

Student Evaluation. *English 33 Diploma Examination Marker's Manual January 1986*. Edmonton, Alberta: Alberta Education, 1986b.

Student Evaluation. English 33 Diploma Examination Marker's Manual June 1986. Edmonton, Alberta: Alberta Education, 1986c.

Swartz, Richard, and Douglas R. Whitney. "The Relationship Between Scores on the GED Writing Skills Test and On Direct Measures of Writing." GED Research Studies 6 (June 1985): 4-12.

Sweedler-Brown, Carol O. "The Influence of Training and Experience on Holistic Evaluations." English Journal 5 (September 1985): 49-55.

Vernon, P. E., and G. D. Millican. "A Further Study of the Reliability of English Essays." The British Journal of Statistical Psychology 7 (November 1954): 65-73.

Weidemann, C. C. "Scoring the Essay Test." Journal of Higher Education. 12 (1941): 41-44.

Willing, M. H. Willing Scale for Measuring Written Composition. Bloomington, Illinois: Public School Publishing Company, 1922.

Wiseman, Stephen. "Symposium: The Use of Essays in Selection at 11+." The British Journal of Educational Psychology 26 (1956): 172-179.

APPENDIX A. SCORING DESCRIPTORS--ENGLISH 33

SCORING DESCRIPTORS - ENGLISH 33

Section I: Personal Response to Literature

Thought and Detail

5 - PROFICIENT: A situation or concept has been effectively established. Precise details about actions and/or characters are deliberately chosen and add clarity. Retrospection or circumspection, whether explicit or implicit, is insightful.

4 - CAPABLE: A situation or concept has been appropriately established. Specific details about actions and/or characters are well-defined and plausible. Retrospection or circumspection, whether explicit or implicit, is thoughtful.

3 - ADEQUATE: A situation or concept has been clearly established. Details about actions and/or characters are clear and purposeful but tend to be general. Retrospection or circumspection, whether explicit or implicit, is conventional.

2 - LIMITED: A situation or concept has been vaguely delineated. Details about actions and/or characters are generalized and are inappropriate and/or haphazard. Retrospection or circumspection is obscure or uncertain.

1 - POOR: An inappropriate or incomprehensible situation or concept has been presented. Details about actions and/or characters are irrelevant or absent. Retrospection or circumspection is not present or is confusing.

INS - INSUFFICIENT: There is no evidence of a discernible attempt to respond to the assignment as stated, or the writing is so deficient in length that it is not possible to assess thought and detail.

Organization

5 - PROFICIENT: The beginning creates interest and/or promotes further reading. The ideas and situations are developed by sentences and paragraphs that flow smoothly and coherently to an appropriate and effective conclusion.

4 - CAPABLE: An effective beginning is presented. The ideas and situations are developed by sentences and paragraphs that are coherently related. The conclusion is appropriate.

3 - ADEQUATE: The beginning, development, and conclusion are functional. Sentences and paragraphs are generally related, but coherence falters on occasion.

2 - LIMITED: The beginning and/or conclusion are non-functional. Relationships among sentences and between paragraphs are frequently unclear.

1 - POOR: The introduction and/or conclusion, if present, are obscure. Sentences and paragraphs are not coherently related.

Matters of Choice

5 - PROFICIENT: Choices made by the writer are usually effective. Many words are chosen for effect and most are correctly used. Many sentences are deliberately structured for effect.

4 - CAPABLE: Choices made by the writer are often effective. Some words are chosen for effect, and most are correctly used. On occasion, a sentence is deliberately structured for effect.

3 - ADEQUATE: Choices made by the writer are occasionally effective. Most words and structures are used correctly, but attempts to structure sentences for effect are rare.

2 - LIMITED: Choices made by the writer are often ineffective. Many words and structures are misused. No attempt is made to use sentence structure for effect.

1 - POOR: Choices made by the writer are usually ineffective. Words and structures are misused to such an extent that clarity suffers.

Matters of Convention

5 - PROFICIENT: The writing is essentially free from errors in spelling, punctuation, and grammar. Errors that are present do not reduce the clarity of communication.

4 - CAPABLE: Few errors in spelling, punctuation, and grammar are present. These errors sometimes reduce but do not impede the clarity of communication.

3 - ADEQUATE: Occasional errors in spelling, punctuation, and grammar are present. Some of these errors impede the clarity of communication.

2 - LIMITED: Frequent errors in spelling, punctuation, and grammar are present. Many of these errors impede the clarity of communication.

1 - POOR: Numerous errors in spelling, punctuation, and grammar are both noticeable and jarring. Most of these errors severely impede the clarity of communication.

Section II: Functional Writing

Thought and Detail

5 - PROFICIENT: A clear and effective awareness of audience is demonstrated. Significant information is presented, and this information is enhanced by precise and appropriate details that effectively fulfil the purpose.

4 - CAPABLE: A clear awareness of audience is demonstrated. Sufficient information is presented, and this information is substantiated by appropriate details that efficiently fulfil the purpose.

3 - ADEQUATE: A recognition of audience is demonstrated. Sufficient information is presented, and this information is supported by enough detail to fulfil the purpose.

2 - LIMITED: Recognition of audience is demonstrated but is not sustained. Essential information may be missing. Supporting details are scant and haphazard so that the purpose is only partially fulfilled.

1 - POOR: Only a vague recognition of audience is demonstrated. Essential information and supporting details are lacking. The purpose is not fulfilled.

INS - INSUFFICIENT: There is no evidence of a discernible attempt to respond to the assignment as stated, or the writing is so deficient in length that it is not possible to assess thought and detail.

Organization

5 - PROFICIENT: A clearly focused and effective statement of topic or function is presented. Details are coherently related to the topic or function and are developed in sentences that flow smoothly to an effective and appropriate request, statement of expectation, or other conclusion.

4 - CAPABLE: A clear and obvious statement of topic or function is presented. Details are clearly

related to the topic or function. The request, statement or expectation, or other conclusion is appropriate.

3 - ADEQUATE: A generally clear statement of topic or function is presented. Details are organized so that a general relationship to the topic or function is maintained, but coherence falters on occasion. The request, statement of expectation, or other conclusion is functional.

2 - LIMITED: A vaguely focused statement of topic or function is presented. Details are vaguely related to the topic or function. The request, statement of expectation, or other conclusion is unclear.

1 - POOR: An obscure statement of topic or function is presented. The relationship between chosen details and the topic or function of the writing is obscure. The request, statement of expectation, or other conclusion is absent or inappropriate.

Writing Skills

5 - PROFICIENT: The selection and use of words and structures is usually effective. Errors in spelling, punctuation, and grammar that are present do not reduce the clarity of communication.

4 - CAPABLE: The selection and use of words and structures is often effective. Errors in spelling, punctuation, and grammar sometimes reduce but do not impede the clarity of communication.

3 - ADEQUATE: The selection and use of words and structures is occasionally effective. Errors in spelling, punctuation, and grammar occasionally impede the clarity of communication.

2 - LIMITED: The selection and use of words and structures is often ineffective. Errors in spelling, punctuation, and grammar often impede the clarity of communication.

1 - POOR: The selection and use of words and structures is usually ineffective. Errors in spelling, punctuation, and grammar severely impede the clarity of communication.

Section III: Response to Visual Communication-Question 1

Thought and Detail

5 - PROFICIENT: Given the situation present in the comic strip, cartoon, or photograph, the writer has effectively and consistently assumed an appropriate role. Feelings are effectively established. Precise details are deliberately chosen for support and/or illustration of the feelings presented.

4 - CAPABLE: Given the situation present in the comic strip, cartoon, or photograph, the writer has clearly assumed an appropriate role. Feelings are clearly established. Specific details used for support and/or illustration are well-defined and accurate.

3 - ADEQUATE: Given the situation present in the comic strip, cartoon, or photograph, the writer has assumed a plausible role. Feelings are explained. Details used for support and/or illustration tend to be generalized.

2 - LIMITED: Given the situation present in the comic strip, cartoon, or photograph, the writer has attempted to assume a role but has not sustained this attempt. Expression of feelings is inconsistent or inappropriate. Details used for support are inappropriate and/or unclear.

1 - POOR: Given the situation present in the comic strip, cartoon, or photograph, the writer has not assumed a plausible role. Feelings are vaquely [sic] delineated. Details are irrelevent [sic], inaccurate, or absent.

INS - INSUFFICIENT: There is no evidence of a discernible attempt to respond to the question asked, or the writing is so deficient in length that it is not possible to assess thought and detail.

Writing Skills

5 - PROFICIENT: The selection and use of words and structures is usually effective. Errors in spelling, punctuation, and grammar do not reduce the clarity of communication.

4 - CAPABLE: The selection and use of words and structures is often effective. Errors in spelling, punctuation, and grammar sometimes reduce but do not impede the clarity of communication.

3 - ADEQUATE: The selection and use of words and structures is occasionally effective. Errors in spelling, punctuation, and grammar occasionally impede the clarity of communication.

2 - LIMITED: The selection and use of words and structures is often ineffective. Errors in spelling, punctuation, and grammar often impede the clarity of communication.

1 - POOR: The selection and use of words and structures is usually ineffective. Errors in spelling, punctuation, and grammar severely impede the clarity of communication.

Section III: Response to Visual Communication-Question 2

Thought and Detail

5 - PROFICIENT: Interpretation of the comic strip, cartoon, or photograph is insightful and appropriate and is in the form of an effective generalized idea or theme. Precise details are deliberately chosen for support, and add clarity.

4 - CAPABLE: Interpretation of the comic strip, cartoon, or photograph is appropriate and is in the form of a generalized idea or theme. Specific details used for support are well-defined and accurate.

3 - ADEQUATE: Interpretation of the comic strip, cartoon, or photograph is conventional and may be in the form of a maxim or moral. Details used for support are clear but tend to be generalized.

2 - LIMITED: Interpretation of the comic strip, cartoon, or photograph is vague and uncertain and concentrates on a particular detail rather than the cartoon or photograph as a whole. Details used for support are inappropriate and/or unclear.

1 - POOR: Interpretation of the comic strip, cartoon, or photograph is inappropriate, implausible, or incomprehensible. Details are irrelevent [sic], inaccurate, or absent.

INS - INSUFFICIENT: There is no evidence of a discernible attempt to respond to the question asked, or the writing is so deficient in length that it is not possible to assess thought and detail.

Writing Skills

5 - PROFICIENT: The selection and use of words and structures is usually effective. Errors in spelling, punctuation, and grammar that are present do not reduce the clarity of communication.

4 - CAPABLE: The selection and use of words and structures is often effective. Errors in spelling, punctuation, and grammar sometimes reduce but do not impede the clarity of communication.

3 - ADEQUATE: The selection and use of words and structures is occasionally effective. Errors in spelling, punctuation, and grammar occasionally impede the clarity of communication.

2 - LIMITED: The selection and use of words and structures is often ineffective. Errors in spelling, punctuation, and grammar often impede the clarity of communication.

1 - POOR: The selection and use of words and structures is usually ineffective. Errors in spelling, punctuation, and grammar severely impede the clarity of communication (Student Evaluation, 1986b, pp. 32-37).

APPENDIX B. RELIABILITY REVIEW PAPERS - JANUARY, AND JUNE, 1986

Section I: Personal Response to Literature - R.R. #5 - January, 1986
Appendix B Part 1

REVISED WORK

People receive gifts every day. Everyone has a different meaning of a gift. Some people receive a gift as a token of appreciation for something they have done for someone. Sometimes gifts are given at weddings or christmas parties. What gift is the greatest gift in life? Most people don't think of it as a gift. I think the greatest gift of life is to be talented in more than one field. Sometimes we think of money or expensive clothes as great gifts. But is it really the article or the Tought behind the gift?

Some people are very gifted in doing there carpentry work around the home. Can you imagine the expense you can save by building your own rech room in the basement. Or else you can build your garage. The main reason people do not have some of these projects done because they don't know how. Mind you, sometimes they don't even want to try. I have done some research into this field and the findings are that for every dollar we spend on building a garage, sixty cents goes towards labour and forty cents of that dollar goes towards materials. So now we know how people can save money doing certain projects. Also people in this field, gifted to do projects like this should be proud of themselves.

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

as I am. I have done my own basement and built my own garage.

Other people have taken courses in day school or night classes and know how to fix there cars. There is a great expense in owning a car. So if you can do some of your own work, you can save a great deal of money and inconvenience. Suppose you need an oil change in the car and have plans to go to Jasper for the weekend. Most of the times you need your car for work so you can't leave your car at the shop to be repaired. Believe me, I work on my own cars and I really have more trust in them because I know the car inside out. If I go for a trip, my trip is not spoiled due to worries of the faith of my old car. I know it will take me where ever I want to go.

Then many people are gifted in being calm and getting along with people. How many times have you been at a party and the same peopl get into arguments? Some times I think people argue for attention. But isn't nice to go to a park or where ever and have some one come up to you and have a friendly conversation. This really makes me relaxed and appriciate living when I see people smiling, laughing, and joking around. We live only once so it is really nice being around people that can get along with everyone. Everyone one has

There is additional space for Revised Work on page 9.

REVISED WORK

a different opinion of being gifted in this but my personal belief is if you work hard enough to get along with people, you get gifted in your ways.

Being gifted in certian fields is a personal satisfaction for my self. But to be gifted in something, you have to work hard at it. It doesn't come on its own. If you educate yourself in certian fields, you end up being gifted in it. The personal satisfaction I get when I stop on the side of the road to help a helpless motorist is unexplainable. When the person thanks you and the smile you receive from him or her is worth more than money can buy. I think most people forget the great gift of life they have. This gift is usually taken for granted and if people do have it, a majority of people either do not believe they have it or recognize they have it. Believe it, the gift to get along with around, repair vehicles, do carpentry work, or even someone stopping to talk to me is the greatest gift of life to me. Its the Thought behind the gift not the article that is a true gift.

Nalwen Composite High School

P.O. Box 203, Nalwen, Alberta. T5J 2R2

Mr John Kennedy
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr Kennedy:

On June 21, 1986, at nine hundred hours to sixteen hundred hours, Nalwen Composite High School is putting on a Carreer Day for the students. There has been a great interest in the field you specialize. I have been appointed to contact you to see if you would come to the school and give some lectures on this special field.

We have at least thirty students interested in the paramedic's work life. I am sure there will be many questions to answer as this is a new field of work. We would appreciate if you could bring pamphlets or any material you can organizing for this lecture. Also if you would like to set up a work shop in the gymnasium, we will supply proper amount of space for your equipment.

Some of the questions to be answered from my fellow students and also myself are as follows. What are some of the requirements to get into paramedic's? Is there a future in this field? What is the work force like? Is there a great amount of stress in this field? Is this program recognized accross Canada or just in Alberta?

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

There are many unanswered questions from student at this moment.

I would appreciate an answer from you if you will acept our invitation to Carreer Day or not as soon as possible. If it is not possible, please inform me who I can get in touch with to assess us and help us make Carrier Day a success.

I will be looking foreward to hearing from you.

Yours Truly,
Pat Jones

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.

REVISED WORK

Did you ever hit your finger with a hammer? I think just about everyone has at one time of there life? It usually happens when you are in a hurry. Patience in the life-saving technique. Relax, and the project will be complete. I know every-time a person is in a hurry, somethink slips and there goes your day. For instance this man, as it has happened to me, being in a hurry because you are just about finished, and have plans to go out for supper, hit your finger at the most inconvenient time. It's mostly to carelessness. You've just about finished the project and bang you hit the finger. I can see my own expression on my face when it first happened. First of all the pain, of the hammer coming down on your finger, then the sight of your finger. At this moment you don't know if to cry or throw the hammer at the first think that moves. Then you finally let out a holler and let the tears flow from your eyes. At the same time you say to your self, slow down and the project will be completed safely if you are careful. So have patience or pain may strike you at any moment.

Section III: Response to Visual Communication – Question 2

2. **What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.**

REVISED WORK

In both selections we have an area in a residual area. First of all we have people building homes for families to live in. Then certain residual areas are complete and we have homes lined upon both side of the street. We have families living in these homes and raising children. Some of the chores for a young boy is to cut the grass on his lawn. As life goes by, the boy grows up to be a man, he still has responsibilite. Some of these responsibilitie are to take care of his dwellings and part of the job is to trim your lawn every once in a while. During his life time to be a responsible person, he has to get a job to support his family. Some of the ways are to go out and work. For example this person is a carpenter and builds houses for future families. After a days work, he comes home and cuts the grass on his lawn. These selection represent the life style of a boy to man. The antennas on the house show us that many people watch television so it is the life style of a home not of a grocery store.

Section I: Personal Response to Literature - R.R. #2 - January, 1986
Appendix B Part 2

REVISED WORK

Dear Diary Aug 12, 1985

Today I was told a true story about how this old man who was on his death bed when his grandson who was 5years old came into roon and asked "Why he had to die? The Grandfather said, I really do not want to die but it comes nautrally. The Grandson asked what is nautrally. The Grandfather said, nautrally is like the changing of seasons in a year. The spring time the year is young and full of life. In the summer time everything is beautiful in every way. In the autumn every thing ~~strats~~ starts to die off or go away. In the Winter time every thing is dead.

life is like the changing of seasons too because when your are young you are full of life and happiness. When you get a little bit older every thing is full of happiness. When you get alot to older the road to happiness comes to an slow end. When you are ready to die you feel a sense of happiness like you had when you where younger.

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

Dear Diary Aug 13, 1985

The story I told you yesterday gave me confidence in living a full ~~li~~ and happy life.

There is additional space for Revised Work on page 9.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta: T5J 2R2

Mr. Pat Jones
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr. John Smith:

Iam a student at Nalwen Composite High School and I would like to invite you to give a speash at our Career Day on June 21, 1986 about your occupation.

Here are some of the questions that you will probably be answering, How is your business doing, What are your hours, Do you like your occupation, has it been successful since you started, What is your salkry.

I would like to Know what you had to have to get into your occupation because I am very intrested in your occupation.

Truthly yours,
Pat Jones
Pat Jones

There is additional space for Revised Work on pages 15 and 17.

Section III: Response to Visual Communication – Question 1

1. **Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.**

REVISED WORK

I would feel discusted because everyone built the same style of house. Everyone is doing the samething. Everybody looks alike and they all live in the same neighborhood.

Section III: Response to Visual Communication – Question 2

2. What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.

REVISED WORK

The main Idea of the cartoons is that now-a days the people can not afford big fancy houses so the have to te build simple houses and built it themselfs. One detail in Section 2 it show all the men on the roofs are having pain problems withe their fingers The second detail in section 1 the tittle tells us that the same person built all the houses.

Section I: Personal Response to Literature - R.R. #4 - January, 1986
Appendix B - Part 3

REVISED WORK

I was working at the Fantasyland Petting zoo in West Edmonton Mall, my first real job besides babbysitting. The Saturday afternoon was just dragging unit my boss decided to liven it up for me.

"I'm sick and tired of you kicking the kids ~~out~~ out of here. Do you favor the animals that much that you won't let anyone touch them? This is the Petting zoo Tony, you ~~half~~ to have to let the kids pet those stupid animals. Look, all you have to do is keep it clean in ~~[illegible]~~ here. That all you have to do for these creatures. I don't want to see one more ~~kid~~ ~~kid~~ kid kicked out of here, ~~you~~ comprehend?' he belched at me.

"Know it doesn't comprehend I." I said with a crackle in my voice. "I won't stand for no kid to come in here and Terrorigh these animal by Pulling the tail, ~~eye~~ ers, poking there eyes out

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

or anything else they get off on. These animals can't just get up and leave here, ~~there~~ they're stuck here, and they know it. This is there home right now and they do intend to get grumpy with kids who irritate them and there property. They'll bite or kick those kids without hesitation and if they, ~~the kid~~ those ~~#~~ brats bug them, there out. I looked him straight in the eye. ~~He~~ Nijiael turned away and started to walk off releasing, "You heard me Tony, its your last chance."

Then, just after my boss, Nassei left, Tinia came up to me. She was a regular ~~ads~~ customer here at the zoo, and ~~I~~ I'm fou favorite. She was a sweet little thing around ~~5~~ ~~on~~ five yeas old with golden blond hair, blue eyes and a million dollar smile. She was wearing a cute pink jumper that really made her eyes sparkle. She looked up at me and said "[illegible] Tommy, I don't

There is additional space for Revised Work on page 9.

REVISED WORK

mind if you kid those mean kids out of here. You take good care of the animals!" Then she handed me a piece of paper and said "Thank-you for take such good care of the animal, I love ya." and she ran off.

I opened the paper. It was a picture of me perticting the animals and Tinia giveng me a hug, with some printing bellow. "You are a GREAT PERSON TONY I LOVE YA." My heart melted. I felt really touched by this. It was the nicest thing anyone ever gave me. and I'll never forget it.

As I look ~~bag~~ back at what the little girl gave me I reallize it wasn't just a gift, it was a bundle of love. Her love.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta. T5J 2R2

Mr. Ein Stain
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr. Ein stain:

Our ~~shcool~~ school Nalwen Composite High school is having a "career day" on June 21rst/86. I have had a great intrest in your work for quite a few years. Gemology, seems to me, to be an interesting and a good career ~~to~~ to pick and I was wandering if you would be interested in ~~or~~ sharring a little infarbmation about your career with ~~our~~ me and my shcool.

I have always wondered how long and how much effort you have to put into this career because there is little information on ~~G~~ Gemology. ~~thing~~ Questions like what courses did you have to take and for how long would I be taking them for, Is there mainy jobs for a Gemologist here

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

as any where else, is it a good career to get into, what is the pay like, can we survive on it and exactly what all do you do, being a gemologist. These are just a few question which I am curious about. I have a question sheet attached to the back of this along with one of the posters which have being placed through our school. If you could think of anything else we should know about your career ~~you~~ that would be great. The more we know, the better it is.

The reason I picked you to write to is because, I heard you are excelent at your work and you believe in helping others. It would be the biggest help if you would show up at our "Career day" program. Please consider it in your busy time table ~~and~~ You can reach me at the school by ~~f~~ phoning

There is additional space for Revised Work on page 17.

REVISED WORK

487-5879 and asking for ithim to
page me, Pat Johns. Hope to here from
you soon.

Thanks again,
Pat Johns

Section III: Response to Visual Communication – Question 1

1. Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.

REVISED WORK

~~If~~ If I was the first man in the second Cartoon I would be feeling a tormendous amount of pain. Hitting my ~~hand~~ finger with a hammer is not my idea a pleasure. Then after I realize that it was my own stupidity for being in all this pain, ~~a~~ I would be mad. Knowing that you can not get mad at yourself and blame yourself for your own stupidity, I would get mad at the close person to me. ~~Then~~ ~~[illegible]~~ After getting mad at this person I feel frustrated because he got mad back at me and he had no reason to. I would be able to taste the gunpowder that I would like to spit out at him, but instead I would walk away. After the pain was gone I ~~felt~~ would feel really guilty for screaming at innocent people.

2. What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.

REVISED WORK

~~People~~ I believe the main idea that is common in both the selection it that niebours, freind or even straigers ~~cop~~ follow or copy others people foot steps. If one person does one thing that looks proper another person will be ~~sh~~ sure to try it.

When the man in the ~~fir~~ 2nd ~~of~~ Selection ~~[illegible]~~ is building his house everyone else diside to build one to because they don't want ~~every one~~ ~~else~~ that one person to have something ~~othe~~ which he doesn't have.

REVISED WORK

This incident takes place in a small town called Markham Bay and its population is about four hundred people. The young boy that is in this story is ten years old, his name is Jonah. The town of Markham Bay has always been an soapstone carving place for many years, and Jonah started up carving because he felt it was used to interpret your feeling's into stone no matter how it looked it would be liked. After he had been carving for awhile he noticed his parents tenth anniversary was in a week. And Jonah said to himself that he has to make a carving for his parents because it is an important occassion, they also deserved something in return from their son. So Jonah gathered up all his tools, his soapstone and got ready for the day he would carve some figure. Jonah had just started

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

carving very young and he wasn't a professional carver yet like many people in Markham Bay. The next day he sat down and started chipping away at the stone. He had to sharpen his axe every once in a while. His imagination and thought about his parents lead him to carve two people holding hands, but it was filed down to more detail. He took his files, chisels, grinder to work and later sanded it down to the last scratch on it. After he'd finished chipping, filing, sanding he waxed it and made it really shiny after all this he wrote on the base and it said "Happy tenth anniversary." And he wrote his name on the bottom of the carving.

When the actual day came he gave his carving to his mother and she asked where it was from and

There is additional space for Revised Work on page 9.

REVISED WORK

Jonah said he made it his mother almost bursted out crying. His father came and liked it very much and thank Jonah. Jonah left his parents alone and wished them a happy anniversary Jonah's Parents were talking how their son is going to become a Caring person and at the same time they were please with him that he has learned to carve on his own. They were shocked of how independant their son was at that age.

This incident has affected me alot of ways, it has showed me how to care for people, how to share with people. And how some occasions that happen around a family can make some differences for later days, or years. This incident becomes a happier story after Jonah showed his appreciation of having his parents around to help him own throughout his life.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta. T5J 2R2

Mr. Ralph LaPierre
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr. LaPierre:

Hello, my name is Pat Jones and I am a member of our school student council and I have been asked to invite a speaker for our annual careers day that will take place on the twenty-first of this month. The school had a survey of what kind of jobs would interest the students. And the R.C.M.P. were favoured, also there is a considerable amount of interest with the different programs the force has to offer. And the student body thinks that a job with the R.C.M.P. would involve quite a bit with the public. The students are eager to ask some questions.

If you would be able to come, such things as pamphlets

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

slides would be very interesting to look at. If you or anyone else is unable to come could you kindly call me at the school. On behalf of the students of Nalwen Composite High School and myself we greatly appreciate you consideration and cooperation.

Yours Sincerely,

Students and

Pat Jones.

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.

REVISED WORK

If I was a character from selection two I would be in pain because if I hit my finger with a blow from a hammer it would hurt, I also have hit myself with a hammer several times and know how it feels.

I would Also be wondering how we hit our fingers all at the same time, rarely do you see a coinsidence with six men working on a house hit their fingis with a hammer all at once and saying "aaaL!" all at the same time.

The Six men are not working with good safety equipment, and the fingis might represent the situation. Because men in the trades don't usually just sit at the roof and not have a ladder to get down.

Section III: Response to Visual Communication – Question 2

2. What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.

REVISED WORK

The main idea that is common to both selections is that some men do the same kind of work on the same type of building. Selection one has these small buildings almost identical to each other but with men working at different areas. And selection two six men, six houses, six hammers, and six fingers in pain and that is repeating itselves all the time.

It can also have a different idea common to the two selections. Both selections gave out a message with the similar line up of house, men with similar tools it is like anything else that has an end. Men work really hard all their lives and it has to end somehow and the fingers might represent something that ends along with pain. And many people getting interested in going to the trades field of jobs.

REVISED WORK

At Christmas night all the sensations and feelin .. are extraneous the main and heart go up and down with happiness and sadness, tears and smiles and for that reasons when Carlos, my husband, give me a big red wrapped box with a tremendous yellow ribbon I almost jump up ~~out~~ ~~the~~ to the ceiling and fell downstairs in my mother in law house..

At that time, Everybody in the room were looking at me, I felt on my hands the expectation of twenty four pair of eyes and I could even more one of my dummy fingers. Finally, Carlos came on my help, he puted his arms around me and kissed my cheek while I started to open my gift.-

Take me at least five minutes unwrapped the big box and peek inside her but my happiness was transformering in madness because inside the box were paper, paper and a candy bag everybody started laughing and my madness melted

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

on tears. My husband was laughing with them and I wished to have a gun for shooting him, for a few minutes all I could see through my tears were smiles but when Carlos asked me if I wanted a candy for sweet my life all the sound I could hear was the candy-paper while my hand was inside the candy bag.- I touch something inside than were not candies, slowly I pulled out and looked, I couldn't talk, I could cry, I was astounding because on my hand I had the most beautiful diamond ring I never saw.-

Long time has past since this incident but always I remember what my husband said to me at that time "You always make a judgment before discerning or compairing"

There is additional space for Revised Work on page 9.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta. T5J 2R2

Mr. Félix Lang Dr.
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr. Lang:

In June 21, 1986 from 9^{00} A.M. to 4^{00} P.M our school plans to have a Career Information day. After an extensive questionnaire about future occupations most of 46% of our students would like to be a doctors.-

We know you have a bussy proffesional life and in the date mentioned before you will be in a good deserved holiday but on the other hand, we know about your interest in young people who like medical career and how you enjoy teaching them about this interesting and fascinating but hard life of been a Medical Doctor.-

Well Dr. Lang, you decide and let us know the time is more convenient for you. If is possible two days before the

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

event, in that case we can write your prestigious name on the pamphlets correspondent to medical career..

Waiting for your answers with big expectations.

Yours Truly

Pat Jones

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. **Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.**

REVISED WORK

Describe feelings is really hard most hard is if I have to put myself like one characters of a cartoon magazine. Anyway I have to decide wich ~~stra~~ characters and I prefer selection one from the National Film Board of Canada called 'The House That Jack Biult"

In the case I were Jack a feel so stupid because all I can'or have to do is cutting the grass around the house and that around mean outside the house, in that case why I build the house if I can not be inside enjoy her?

Section III: Response to Visual Communication – Question 2

2. **What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.**

REVISED WORK

Everybody dream about own a house, everybody makes plans for the time they live in an own house but when the time is coming they can have two attitudes.

One attitud is restraining ~~your~~ their natural personality and take the things so quint ~~on~~ while their are the slaves of the house like cartoon #1

And the second attitude is rebealing against the house's slavery and yelling out what they feell about of be an owner of the house.

REVISED WORK

My sister and I had been taking music lessons for almost two years. Karen taking guitar lessons and I taking accordian

It was nearing Christmas time and [illegible] knew for certain that my parents were getting Karen a six string Guitar. I had no idea what I would be getting, although I figured that if Karen was getting a guitar, I would be getting an accordian. I wanted an accordian so bad.

One evening as Karen and I laid in bed making wishes for Christmas gifts, Karen had asked me if I knew what Mom and Dad were giving her. I had told her that I would tell her in exchanged that she would let me know what I was getting.

Karen had said, that I was going to get a sterio for Christmas.

I was heart broken knowing I wasn't getting my accordian.

I told Karen about her guitar and she was tickled pink.

Christmas eve finally rolled around, the Christmas gifts were under the ~~exchanged~~ tree except Karens guitar and my sterio

We had always exchanged our gifts on Christmas eve.

My parents went outside for a few moments and walked back in the house with two large boxes. One definitly was Karen's guitar as you could tell by the shape and size of the case.

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

Karen was sitting on the couch waiting impatiently to try out her new Guitar.

They put my big gift beside me. I knew I had to try to look excited although I had felt that I had been cheated out of something that I really wanted.

I started taking my time opening the gift, and once the Christmas wrapping was off there was a cardboard box. The box was taped closed so I took the my fingernail and sliced the tape. I opened the lid and inside was a grey and black case.

My first thought was that my parents had bought me one of those stupid old fashioned stereos like they had in elementary school.

I took the case out of the box. I laid it upright to open it; I clicked the fasteners, opened up the lid; There in front of me was an orange crush velvet lined case. There was a piece of the same material over a large object, I lifted the material and there it was; a 120 bass Rossini Pearl Accordian. I was totally shocked. I was so happy tears began to swell in my eyes. I caressed my new accordian and then ran over

There is additional space for Revised Work on page 9.

REVISED WORK

to my parents and gave them the biggest hug and kiss I had ever given.

I had noticed Karen sitting on the sofa giggling. She said in an uncontrolable ~~lafter~~ laughter "Gotcha". She had known all along what I was getting, but she didn't want to tell me.

I had felt bad for telling her what she was to be getting, but she certainly didn't seem to mind. Karen was just as happy as I was.

The rest of the evening Karen and I took turns showing off how well we could play on our new instruments.

My accordian was the best ~~gif~~ gift ever.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta. T5J 2R2

Mr. Barns
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr. Barns:

Nalwen Composite High School is having a Career Day for the students, in order for them to help plan their future.

Ms. Barns, your expert reputation has much suade our decision in asking you to be our Spokesmen.

Our Board of Directors would be Honered if you would accept.

You would be required only to give a lecture on various types of Courses our Colleges and Universities are offering. The main interest of most of our pupils is in the Business courses, so we could use as much of your knowledge as possible.

You would be required to explain to them how they would go about registering in to a University or College.

cont ->

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

As well we would like to ask you to explain how they would go about requiring for a Government Grant or Student financing.

Our Career Day is on January 17th/85. We would appraciate to hear from you by the 9th of January if you will accept or decline our offer.

Sincerily Yours,

Pat Janes

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.

REVISED WORK

I chose selection or Cartoon number two, as I would feel a little embarresed being a carpenter or roofer and hitting my thumb with a hammer, as I see it these men look inexpersencied with the work that they are doing.

Section III: Response to Visual Communication – Question 2

2. What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.

REVISED WORK

I feel the main idea between the two of the cartoons is the maintenance of taking care of your property and building.

In cartoon one the man obvouily is taking care of the appearence of his lawn, and in cartoon number two the men are taking the care of shingleing there house to keep them from leaking in the rain and seasonal errosion.

Section I: Personal Response to Literature - R.R. #3 - January, 1986

Appendix B Part 7

REVISED WORK

It was almost time for my mothers birthday. At that time I was just a young boy, seven or eight years old. We lived in a rural area; our small farm provided little more than what we needed for sustenance so there were no allowances and seldom any money for gifts. Although there were ways for enterprising children to make money, picking bottles or doing odd jobs for neighbors, the older kids would leave little for us younger ones.

I was in a quandary. Not having a gift for mom on her special day did not seem right. The sum total of my resources consisted of: a few pennies, a jacknife, a whistle, the hide of a mouse, and a toy pistol with a broken handle. Treasures to me but nothing having any commercial value.

At that time children were "seen and not heard" and did not dare seek handouts from adults. Money was scarce and if we could not earn our own, we had to do without. I done the best I could but there were no bottles in the ditches and my subtle solicitations for gainful employment went unheeded.

I was broke!

continued.

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

The closer it got to moms birthday, the worse I felt. I would have nothing to give this warm wonderful person who deserved everything. I looked again to my possessions, nothing of interest to a mother, nothing at all.

The day arrived. I left the house early that morning. I was feeling so bad that I decided to spend the day holed up at my hideout up by the lumber pile. Playing was no fun, even my dog sensed my mood and went off seeking more lively companionship. I moped around the fort until my stomach called, then went to the house for a feed and quickly left. I'm sure mom must have noticed that I wasn't myself, but with the eternal wisdom of all mothers, said nothing at all. As the day wore on I felt worse. A big solid lump filled my chest, and it seemed that nothing could ever make me feel good again.

I knew it was close to supper time and that mom, grandma, and various aunts and neighbors had spent hours over hot stoves preparing a birthday feast. I had no interest in food, and even less in seeing or talking to anyone. Oh the depth

There is additional space for Revised Work on page 9.

REVISED WORK

of the misery I had sunk to. I sat in my fort, feeling about as bad as a human could feel. Then I heard mom calling me for supper. She sounded happy and excited, and ready to get on with the celebration. I could'nt respond, I didn't want to face her. She called again then I could here her walking towards the lumber pile. The footsteps got closer then her face peered through the door. I burst into tears - the damn broke and it all came out. Between sobs and sniffles I managed to tell her how badly I felt. A tear trickled down her eye and soon both of us were crying, but not from sadness.

Mom took my hand and we walked down the hill and into the house. We sat down and then she told the guests that she had just recieved the finest gift a mother could ever hope to get.

In remembering and writing about this incident I have relived it. I cannot recall any time or event since then that has moved me so much.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta. T5J 2R2

MR. Arthur Pendergrass
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

Dear Mr. Pendergrass:

The faculty and students at Nalwen Composite High School will be holding a "Career Day" on June 21, 1986. I would like to extend an invitation to you to appear as a representative and guest speaker on behalf of the resource management field.

Our adgend Agenda allows for a half-hour presentaton from each guest speaker, and we have been granted use of the school audio-visual equipment if you would need them. If you accept this invitation we would appreciate any pamphlets or other written informaton on this field.

Lunch and other refreshments will be served and any assistance you may require will be provided.

Sincerely,

Pat Jones.

There is additional space for Revised Work on pages 15 and 17.

Section III: Response to Visual Communication – Question 1

1. Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.

REVISED WORK

If I were the man in the first selection I would try NOT to feel. There is little or no stimulation in this situation. I am doing a repititious job mowing a similar lawn~~s~~ in a neighborhood where all the houses are the same. Nowhere is there any evidence of thought, feeling, or creativity. Everyone else in the neighborhood looks the same, dresses the same, and even mow their lawns the same way at the same time.

Section III: Response to Visual Communication – Question 2

2. What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.

REVISED WORK

I think that the main idea in both selections is to illustrate how much we are influenced by what goes on around us. In both cartoons we see people that are doing exactly the same thing as the other people. Cartoon 1 shows similar people mowing similar lawns in front of similar houses. Cartoon 2 shows six men working on the same end of the roof of six similar houses, having exactly the same finger hurt at the same time.

REVISED WORK

There is a person whom I tremendously admire. She is a girl I've known for twelve years and her attitude is extraordinarily unexpected.

This girl has a fascinating reputation because of her honesty and encouragement she has to make others feel good about themselves. She has a high degree of willingness to help someone get out of a problem and try to comfort you while she's doing it.

One incident she reveals is her wantingness to help me if I am lost with my feelings. She makes me feel happy inside with a sign of wanting to go on. Even though the problem looks unsolvable, she manages to find a bright side to every situation. She is always there to bring cheerfulness back into my life again.

Her unexpected behavior is unbelievable. She will say things that nobody else can think of. In her ac

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

mind. She never worries about what other people think of her or might say to her. The attitude she has about life is incredible. If something goes to her disadvantage she won't sit around feeling sorry for herself and hope it goes away. Instead, courage of carrying on is always with her. There is always a solution even to the most difficult problems. To give you a sense of hope she'll say, "it probably just wasn't meant to be," or, "at least you tried a know now instead of wondering what might have been."

This young adult's behavior is memorable to me because she's happy, is herself, and never puts on an act. She's regularly out to make people feel good about themselves and have confidence to carry on. She amazes me because of this. To her things are taken as they come, and

There is additional space for Revised Work on page 9.

REVISED WORK

She always has a tendency to be willing to advance to something new. She's is a very special frie. to me, and it makes me very proud to admit her generosity a. unreal personality she has.

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

I have recently gone through some of the situations and have noticed the one disadvantage that is making one section of our city suffer. It is the senior citizens lack of community services. I, myself, am a senior citizen and I think it would be worth everyone's while to try and get more programs going for us. If this means having to take a survey or voting on how the majority feels, then this task should be completed as soon as possible.

Senior citizens have played a very important role in our society, and I feel they are just as important to look out for. For example, we take

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

part in most of the city's events, and we involve ourselves in the society's advantages and disadvantages. I feel that if we are living in the city and carrying on with the situations happening here. we should be thought of as responsible people who at least deserve a chance to see what this experience would be like.

All we are interested in right now, is a chance to speak up and say what we feel. This could turn out to make thousands of people happier, but if this event doesn't work out at least we had a shot at it, and we will know better next time to compete with more positive feelings.

Thank you for taking the time and being considerate enough to imagine how we feel. Could you please give

There is additional space for Revised Work on page 17.

REVISED WORK

further instructions if you decide to consider this? You can always contact me at 486-3105. If this involvement comes through I am looking forward to your dignity.

I really appreciate your thoughtfulness.

Sincerely yours.

Pat Jones.

Section III: Response to Visual Communication – Question 1

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

I think the character who is pretending to be a flower is feeling a sense of acheivement because of her willingness to try. She, at her young age, is probably frightened at having to face all the other children watching her. The character doing the sherade should be very proud to be able facing her fear, and giving a shot at what she is extremely paranoid of. This girl is trying to picture herself as many things that she wants to be Whether it's a an idol, or just someone to look up to, she like all youngsters do, is picturing in her mind what kind of a person she wants to be. This attitude is very healthy, because it helps her to decide and see all the many sides of life. I think she is feeling a sense of overcoming one of the fears she will face when she is older.

Section III: Response to Visual Communication – Question 2

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

The main idea in this cartoon is how all the children have different views on life. They are all thinking of something unique and to their liking which gives them an idea about the real world. Kids are always thought of imagining many things when they are young, it helps their minds to grow to either change their opinions or to feel even stronger about them. This is why children have idols. It's someone they adore and will always look up to, even though it's just a figment of their imagination. This cartoon is just telling to think what we as individuals want to be, not as what someone else wants. for you. Having your own mind and wanting something is what only you can decide and it will be to your advantage in the end.

Appendix B Part 9

REVISED WORK

I'm writing to you about an article I read in the Edmonton Journal about a grinding forward who's name is Chris Nilan and he plays for the Montreal Canadians. In his social life he is very calm yet on the ice he is a screaming bull.

His public reputation is very good yet he is known as a fighter to all his fans. The fans go crazy and start to chant Chris, just because he started fighting with a person on the opposing team. One night in the Montreal Forum, Chris Nilan could have fout one of the Boston Bruins players, but he didn't. Well the fans turned against Chris and they started to boo. The next morning he saw head lines which read Chris Nilan chickens out. Nilan was now known as a big chicken. So he had to prove himself the next

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

time he plays Boston. He got his chance in the Boston Gardens to hammer on Brad Park's head and he did just that. People now loved him, because he is now once again a tough guy for the Montreal Canadians.

Chris Nilan as I knew him was a very tough person, but the night against Boston, he proved that he just didn't care about fighting. He likes hockey for what it was and thats hockey. Brad Park deserved every minute, because he thought he was a big tough guy until Chris hammered on his head. This is one incident I will always remember, because it has a learning expirience in it. ~~[illegible]~~

I'd wish to get some more information on Chris Nilan as a hockey player. Thank you for reading my letter and you can get a

There is additional space for Revised Work on page 9.

REVISED WORK

holed of me at 429-8859 and Thanks again.

Yours Truly
Pat Jones

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

My name is Pat Jones and I would like to bring to your attention about the condition of the community hall in Nalwen.

Last week my team held a hockey banquet in the community hall. The hall looked like a shack, something you would find on a old farm. The rug was all torn in the entrance and there was graffiti on the walls. This was an inconvience, because everyone hated the place and wanted to leave. I feel this was very embarrasing for me, ~~and~~ because I booked a banquet there and second for the little community of Nalwen. The greatest concern

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

to me is the safity of the place. It looks like it might cave in from the inside. We need to take action fast so we can save the hall and the respect of the community.

Some suggestions are we get all the people in the community of Nalwen and start rebuilding the place, or we can hire some professionals to redo the whole place so then we can gain the respect back of the people in the neighborhood of Nalwen. Maybe we should also fire the janitor, because he obviously didn't care for the community hall either.

Thank you for taking time to read my letter and you can contact me anytime and anyday at 426-8787 so we can go over some of my ideas.

Thank You
Pat Jones

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.

REVISED WORK

The young boy who is thinking about the airplane is preparing himself mentally to go up next in front of the class and do his act.

The class is playing shurades, because everyone is thinking about an object or creature. The boy chose a plane probabily, because no one else did. It is easy to do and probabily he wants to be a pilot our he just likes planes.

He wants to be a pilots and the class is enjoying themselves by doing something different and playing shurades is one way.

Section III: Response to Visual Communication – Question 2

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

The main idea being presented is that everyone is enjoying themselves and school.

The teacher is trying to make the children at a young age think school is fun and not always boring. To me school is fun, but I don't remember playing shurades or any games just to make us enjoy school.

School is fun and can get boring, but not to often, because the teachers are nice and you are around your friends all the time.

Section I: Personal Response to Literature

Appendix B Part 10

REVISED WORK

Up to now every one at the Aspin Play ground was, affraid of the king Paul Lloyd. He was seen by the younger uys as being someone you did not without Wont toanoy. They thought of him as being a rabid dog, ready to lash out at them any chance he got. They realy didnt see him the way the rest of us saw him. We the older group, saw him as being a phoney. We could see threw his front. He tried to act as if he was bad leroy Brown the baddest dude in the hole town. But in reality was just trying to look something he was not. Take last summer as an example. He said that he went to the playground one sunny afternoon at Piegion Lake. He said that he was swing so hard that he broke the swing. He ~~said~~ The funnyest thing about it all was, he said that he had ment to and that he stold candy from the Mulhurst store on his way home. He said when he told his dad, ~~that~~ what had happened he started to pat him on his bake. Paul Lloyd was the kind of person who you thought would laugh if a cat got ran over. Well he's not.

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

last week at the play ground paul and I were fatthing for the king of the play ground position. Paul pushed into a tree, When I hit the tree a birds egg fell from it's nest. Paul quickly picked it up with a terified look on his face and ran home. We didin't see paul for a couple of mounts, When all of a sudden there was paul. Paul invited the playground crowd To come oor and see the bird he hatched. You see people aint always as they seam you just have to look hard enough threw them to see the real person inside.

There is additional space for Revised Work on page 9.

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

I recently saw the artisle in the Nalwen News about the Problems that face the comuntis. I agree that something should be done about this emedetly.

I live in the Nalwen Comunity. Nalwen is about 30 miles away from the city center. We are in the city limits but get no snow removal or road matense.

~~[illegible]~~

The roads are geting so bad that I am afraid to drive my new car down them from fear of falling into one of them. This may sound a little ironic but true. Last week I measured a pot hone on pine lane to be 10 feet long, 3 feet wide and 18 inches deep. That is not a pot hole that is a Crater. The roads havent be tuched since I moved in here 13 years ago. That to me seems a long time to not have done them.

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

Some people in my communty have stated to fill some of these crators by them self. We tried to contact the city on this matter and all they would say is that it would be done withine the next 6 years. By the time six years rolls around the craters will have swaled op my house.

Snow removal isnt much better we dont have eny. If it snows no one goes to to work. With the roads being in such rough shape you cant even snow blow them.

When we asked the city about this matter they told us that we were to face away. this has got our comunity in an uprow. We at this point in moment are willing to do eny thing.

I hope something is done about these matters quickly

Sincerely Yours Six

Dot Jones

Dot Jones

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.

REVISED WORK

The charater With the thought about the sail boat is probalywondering how he is going to do it. He looks as if he dosent think that he can do it. Or on the other hand he sees the girls as acting out a sail boat. This cartoon show that not one person has the same thoughts on. what some one is doing. the boy imagening the thinks that what she has to be doing, his because thats what he wants to see. People see what they want to see. If they dont like wa what they see the imagine it.

Section III: Response to Visual Communication – Question 2

2. **What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.**

REVISED WORK

This cartoon showes the difernces ~~pe~~ bettwen people. It shows that not everyones thought are the same. Everyone has a diffent opinoin about everything. The girle is acting ~~of~~ out a flower but one boy sees it as being tree. People arnt always on the same wave lanth as each outher. People wont to think what they wont not what Joe blow wants. People only care ~~wa~~ what pleases them selves. This cartoon only shows it in great detale

Section I: Personal Response to Literature - R.R. #8 - June, 1986

Appendix B Part 11

REVISED WORK

There was a person I used to know in [redacted]. Daryl [redacted], was his name. He was not a bully, but probably was one of the greatest practical jokers of all time. He was in to all the regular stuff like, snake-in-can tricks, whoopee cushions on seats, getting people drenched!

But, in my gang of friends he was not well liked, because you could never trust him. One day, however, I went to the conveniter I saw him buying dog food. Daryl wasn't known to have dog, so I politely asked him if I could see it. When we got to his house I was shocked by what I seen. The house was a mess, the rug had stains everywhere and there was junk everywhere. His mother was very friendly, but reeked like a brewery. Then I followed Daryl into the back where his dog was. I was surprised to see that the dog was in very good health; well fed and very well groomed.

Late that afternoon I went home and discussed this with my mom and she said that Daryl was the product of a bad

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

relationship with his dad. All his mom lived on was the money he sent her.

One day it became well known that Daryl's dog had come down with distemper. The disease if treated early can be cured, but once its advanced, its only a matter of time. Daryl did everything to keep the dog alive. But one day the dog had a seizure and it was dead.

After that Daryl went bad. It seemed like the dog held everything together. He was thirteen when he hotwired his dad's car. At fourteen he stole everything from cars to magazines. One day he was caught selling drugs and put in jail.

The newspaper said he was a kid who never cared about anything. But I know there was one thing he cared for the most, his dog.

I will forget, Daryl [redacted] because of his love for his dog. Something he did that was very out of character. It was his dog that held him together and

There is additional space for Revised Work on page 9.

REVISED WORK

once that was gone, everything fell apart. I wish people like him got more of a chance

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

I would like to bring to the attention of the editor the road conditions in and around our community of Nalwen.

The quality of roads in and around Nalwen is very poor. In the summer the roads have a number of potholes and when it rains well, forget it, because unless you have a four wheel drive your not going anywhere. on these dirt roads.

In the city the potholes are the worst. As I was going to work this morning I must have hit about three (not good for wheel alignment). The potholes makes this town a very dangerous place to drive, this requires immediate attention.

The conditions of our roads in wintertime. is starting to get worse. I recall just this

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

past winter, one storm left people stranded for days because the roads weren't cleared. I have also heard a number of people in the country had trouble getting into town because the roads weren't cleared. These situations should not happen.

I propose that we get these roads fixed up. The first thing we must do is get some asphalt crews and cover up those potholes. Then we should put gravel on the country roads so that they don't wash away went it rains. I think [illegible] would be nice but we don't have it budgeted.

Snow clearance, I think, should be immediate. Soon as a snow storm hits, crews should be out there clearing, if it last two or three days do it every night. But don't let it pile up or cars won't be able to get through.

These three problems require immediate attention ~~so~~ and should be present ~~before~~ to the public.

There is additional space for Revised Work on page 17.

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

The girl in the front of the room who is pretending to be a flower probably feels like she's getting her idea across, when actually, the class gets very different images of what she's doing. Some of the class thinks she's a fish, some a cat and some an airplane. I would wonder if I was really getting my image across, because you never really know what the class really is going to think, because everybody gets a different ideas about what your trying to portray.

Section III: Response to Visual Communication – Question 2

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

The main idea in the cartoon is about thing you do and how others see it. We may try and portray ourselves in many different ways. We may think we're portraying something, but the whole class thinks your something else. In the cartoon the girl thinks she's a flower, but the rest think she's a tree, a cat, an airplane or a boat. I think that in order to get an image across we have to be very specific or other people will get wrong ideas from what we do

REVISED WORK

In a musty, old warehouse, near the waterfront, a man stood against a wall, in a dark three piece suit. He had a dark complexion, and black hair and a moustache. Beside him were three armed men, also in suits and dark glasses. Across the big room, from him, were two men, bound in rope, their hands and feet tied together.

The man was Fernando Sanchez, a wealthy businessman and ring leader for the underworld. The men beside him were his henchmen. And the two men across the room were his enemies. For a long time he stood and stared at them. He was thinking of how much trouble they had caused him and was savoring their present situation. With an evil grin on his face he looked at his lead henchman. "Hit 'im" he said. His henchman grinned, and then looked at the two frightened men across the room.

As he left the building he could the sound of automatic weapons, muffled by silencers,

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

(con't)

and the screams of the two men, muffled by the sounds of trucks, cranes and other machinery of this industrial slum. He stepped into his Cadillac and drove away.

On his way home, he thought of the two men in the warehouse, and then, of other men like them, enemies. For most of the way home he plotted against more of his enemies. It was all a game to him, a game, which he enjoyed very much.

As he got closer to his suburban home, his thoughts changed. They became more peaceful, As he pulled up his driveway he knew that his wife and daughter would be home to greet him.

"I'm home," he said as he walked into the house. Immediately, his wife rushed to greet him. As they embraced, his five year old daughter came running to him, with a drawing in her hand. "Daddy, look what I made today!" she said. "What's this?" he laughed as he picked

There is additional space for Revised Work on page 9.

REVISED WORK

her up and p kissed her. "What a beautiful picture", he said of the simple drawing. He then, walked into the living room to play with his daughter.

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

I must tell you how pleased I was to see someone, finally, address some of the problems in our community. I have been a resident in Nalwen for the past twenty years, and have watched my neighbourhood slowly decaying.

The biggest problem we have is with our youth. There are no recreational facilities or events for them at all. This is leaving them with idle hands, for which they use to no good. Vandalism has gotten so bad around here that I am afraid to go to sleep at night, wondering what I'll find damaged or destroyed in the morning.

I am definitely not the only person on my block who feels this way. We have all

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

gotten sick and ~~tire~~ tired of this situation and want something done about it right away.

When I was a kid we used to always have something to do. Whether it was a dance or sporting event we were kept busy, and none of us ever spent our time breaking people's windows.

I think that people should start getting involved with this community and organize things for our kids.

Sincerely,

Pat Jones

There is additional space for Revised Work on page 17.

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

The girl in the cartoon is enthusiastic about the show that she is putting on. She is trying to communicate the idea of a flower. Almost ignoring the rest of the class that is watching her, she is very involved in her show.

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

The main idea in this cartoon focuses on the difference in the interpretations ~~that~~ people can get from the same message. The (mime) charade that the girl is putting on is being interpreted differently by all of the children. They also associate her actions with different things. Because of this, they are all getting different thoughts.

REVISED WORK

John

Hows it going? You remember my cousin Kevin right? You met him that one year when he stayed with me and you came up for the week. Well we had quite an experience last weekend.

It all started Thursday Morning. Kevin and I decided to go out to the lake for the week. We called Frank, Sue, Sharon, and Debbie and they all were glad to go. So we loaded up my Chevy 4x4 and got the Odysseys on the trailer then left. We got to the lake (Wolf lake) by around 3:00 pm. We had camp set up by 4:30 pm. Then Kevin, Frank, and I decided to take the Odysseys out for a spin to check out the area. You know how Kevin rides, well it was different this time, it was me who crashed ~~t~~ an odyssey.

We had found this great area to ride in. There was dunes and hills everywhere. I was driving along a ~~k~~ path when Frank cut across in front of me. I swerved to miss him and I hit the hill sideways and rolled the odyssey down it blowing my front tire on the way down. Frank apologized after saying he didn't

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

see me coming. There was a gash about 3 inches long in my tire. I guessed It got cut on a sharp rock. Frank had to tow me back to camp.

I didn't know how to fix it and neither did Frank and it was no use asking Kevin because he could not tell the difference between the front tire or the back. Kevin is a good rider but he does not know how to fix them. So we decided to leave it until morning.

In the Morning Kevin said he could fix my tire. Frank and I laughed at him but decided it was worth a try. Well an hour later Kevin came over and said he fixed it. I did not believe him until he showed me the tire. He had it looking like brand new. Frank and the girls just could not believe it but it was true. Talk about your miracle

We were so glad that Kevin fixed the tire we had a celebration in his favour. The tire worked for the rest of the week. I am still using it for that matter.

That is one weekend that I will never forget. It also helped Kevin to [illegible]

There is additional space for Revised Work on page 9.

REVISED WORK

took a Mechanics course and is a great mechanic now. The main reason I will remember that week though is how Kevin fixed that tire and shocked us all and saying that he saw it done on Television.

[redacted]

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

In response to your editorial in the June 9th edition of The Nalwen News I am sending a letter of something that is of concern in my community. We have been to city hall with our problem before and have not had any response to it. The problem is the main road through our Community is cracked and old. It is in rough condition and is hard to drive on. There are a lot of potholes too. The city has tried filling in the potholes but it does not seem to help any. Also the curbs along the main road are too steep for our cars.

The problem could be corrected if the city could repave the main road and redo the curbs along it. It should not cost them that much money to do it either, all they have to do is put a layer of ashphalt on top

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

of the old ashphalt.

There is another problem that is of some concern in our Community. It is that in the winter months the snowplow always seems to miss our street. It gets hard to drive on our street because of all the snow piled up on it. It gets worse when the snow gets packed down because the roads get icy then. I would like to see the snowplow clean our street every once in a while. It would not cost the city anymore money.

Sincerely Yours
Pat Jones

Pat Jones.

There is additional space for Revised Work on page 17.

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

The character I have chosen is the one thinking of the airplane. I think that he is feeling a sense of adventure. The little girl in front of the class is acting like a flower yet the boy is thinking of an airplane. He is sitting at the back and does not look interested in what the girl is doing. He wants to be in an airplane. He wants to be doing something daring. He looks like he is feeling bored because he is not even associating his idea with what she is doing.

Section III: Response to Visual Communication – Question 2

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

a main idea being communicated is that young children have a big imagination. the little girl up front is pretending she is a flower yet all the rest of the class thinks that she is something else. Everyone is thinking something different.

Section I: Personal Response to Literature - R.R. #14 - June, 1986
Appendix B Part 14

REVISED WORK

Dear John,

Do you remember hearing about a boy from our ~~little~~ high school Harry Jones? I wouldn't think you would seeing it was over fifteen years ago. Anyway, he was the alstar athlete, the ~~spoil~~ rich one, the one with everything. Everyone thought he was so strong, nothing could shake him, worry free.

Well, tho we were wrong. When he finished university he married a beautiful actress, Jane Sones I believe it was. Anyway he started his own law ~~[illegible]~~ firm, John and Associates. He had everything a person could want, when last week I read in the morning paper that he had died. He had comitted suicide.

I could not believe my eyes. I would have done anything to be in the position that he was in, and he just through it all away. I can remember always looking up to him even in elementary school.

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

I feel as though he has ripped away ~~my~~ a part of me, I can't describe it. It was like finding out Elvis had died. That was the one thing that was totally unexpected from him. His wife, kids, and friends are shocked, there was no warning at all. ~~[illegible]~~.

I will never forget that morning. It was such a shock to me that I nearly fainted. ~~The~~ The man I have looked up to all of my life, my hero, ended his life for no apparent reason. I guess I will never be the same. The only one thing I would like to know, the one question that has been eating me up inside for the last few days is simply ~~[illegible]~~ WHY?

There is additional space for Revised Work on page 9.

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

My name is Pat Jones, the owner of the 'Do it Yourself' drycleaning center. I've spent my life trying to create a better life for my family and I. We moved to Nalwen, one of the richer areas in town as you know, to better our standard of living. The first month was exellent. We now lived in our dream home. When our family and friends came to see our new home they were all envious of how well we were doing. The house is still beautiful, exept you ~~[illegible]~~ can not see it for the 20 foot pile of garbage on our front lawn. After three months I thought that the sanitation department must be on strike. To my surprise there not. ~~[illegible]~~ I have spent $400,000.00 on a new home that smells and now looks like the city dump. I was shocked to learn that at the end of this

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

month, the sanitation department is realy going to go on strike. I give 50% of my earnings to the government for protection, clean streets, and to get rid of my garbage. I want some action and I want it now.

Since our mayor does not seem to care what is going on, we should remind him that it is the tax payers of this ~~to~~ town that will be voting in only three months time. If that does not work ~~that~~ then more drastic measures are called for. Sir, you and your paper should advertise a 'DUMP ON PARLEMENT DAY' ~~Day~~. On this day our town people will gather up all their garbage and dump it on the lawn of our city hall. It would be a nice gesture don't you think?

Yours truly,

Pat Jones

PAT JONES

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

I don't believe it! I'm sitting here in this boring class-room wondering weather or not I'll be able to live through another day of this nonsence. Just think! I could be fishing, reeling in the biggest bass ever seen to mankind. But no, here I am watching a girl trying to fly. Oh wouldn't it be wonderful just lying back in the sun drinking an ice cold soda. Ah! Thats the life. I wonder, if I slip out ~~[illegible]~~ now, will the ~~if the~~ teacher ~~will~~ notice? I better not. She's caught me the last twenty-seven times.

Section III: Response to Visual Communication – Question 2

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

This cartoon shows that everyone is different, having different ideas, wants, and worries. Even in this class of young students, living in the same ~~[illegible]~~ area, all of them have experienced or wish to experience totally different things. This cartoon also shows that life is a ~~wonderous~~ wondrous and grand thing that is full of fun, exitment, and new adventures. Life does not revolve around only one person, ~~and~~ what your ideas of ~~[illegible]~~ right and wrong are does not ~~might~~ mean that ~~I~~ they are ~~be~~ the only answers.

REVISED WORK

On Christmas holiday. My friend (Stacie Fink) came to vistid me from Merritt B.C. I couldn't believed myself that she was changed so fast. She grown up, had a husband and a son (Paul). Her son was about 12 yrs of age.

It was a beautiful day. Paul and his neighbor went outside and played their toys. They seemed to enjoyed themselve. and gets along well but all of a sudden, they were fighting to each sides. (and I saw them were fighting) until Stacie came out and interfered the fight.

"Now, tell me what had happened out there"? she asked angrily to her son.

" I was was playing a truck, I put it in the sand and which. I was built but Kevin (the neighbor son) said. " I was not allowed to

There is additional space for Revised Work on pages 7 and 9.

REVISED WORK

played with the sand, and I wanted to, so I just played it anyway, and he stopped me, then he hits on my head, and I didn't pay any attention to him, then he hits me again, so I hit him back." Paul Explained.

His mother was believing every single words he had explained. And she angrily went to the neighbor and complained that her son had bitten up from their son. The neighbor was gentle and pleasant apologized to her and their son had been grounded for two weeks.

I couldn't believed what I had heard. It was all the opposite way but the neighbor (the sons was nice and apologized to Paul and Paul's family. I was trying to explained to Stacie but she wouldn't listened or believed it. I was un-expected that she was changing it.

There is additional space for Revised Work on page 9.

REVISED WORK

her attitude and the way she was living

Paul was saying to himself. : "Oh boy!" that was clase, it should be me to be punished. Poor guy (the neighbor's son) was punished for 2 weeks what a shame!

I will never gets rid of my mind to my dear Paul and Stacie Fink. I don't know what happened in their future because Paul was so young and he was a liar, what happened when he grows up. I hope he will changed to good person.

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

I was concerned about the Park Development. I don't understand why the men (the builders) built the park so far away from our society. I have to drive to the park for at least an hour long. Not because I lived far away but because they built it too far away. Even though, I lived far away, I think they should be built it close and more to our community society. That way if we came late to picnic, at least we have a place to sit or to enjoyed ourselve with beautiful day. It's not that I was too complicated. I don't mind at all to drive an hour long to the park,

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

What it make me angry Was when we got there, there weren't any place for us to sit.

So why don't you build more park that are near to the town instead far away from it. I understand that it will made the community society unclean but the leastest you can do was built up so some more chair to the nearest park or play ground. That way we can sit around and enjoyed our nature society.

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

I was chosen. one of the audience sitting on the left with his two hands holding on his chin on the second row. he was thinking that the actor was acting like a tree. For example. his two hands was raise up with a little round with his arm and it just look like a round of the tree top. his body was skinny and it just like a ~~t~~ tree. His two skinny legs were standing a little apart and my ~~tre~~ stem was ali [illegible] - look like that too

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

The one message being communicate in this visual communication cartoon. is people act that it should be. what it like while she was acting but people as the audiense thinks it was difference way For. example. The girl was acting that she was a flower but to the audiences they think with differences thinking. Such as. one of them thought she was acting like a bird, and the other thought she was a tree, a cat, a plane, a horse, a fish or an octopus Every people have different thinking (thought).

Appendix B Part 16

REVISED WORK

March 25-1986

I am now entering day seven of the ~~monotonous~~ adventurous life of a hospital captive! My captors still can't understand why I don't break down and confess my secret malady. Thus, they are ordering more torture for tomorrow. Trying to break my resistance even further, they are denying me food and water after midnight tonight. I'm not sure how long I can take this punishment. And the saga continues.

On another note - Mom and the Monster came in today.

¶ Considering my Justin's behavior today, I'm not sure if I can call him Monster anymore! My darling six year old brother didn't have a temper tantrum, destroy my room, pull my hair, swing off my I.V. or even bite me. Instead he came up & sat quietly on the bed and gave me a

There is additional space for Revised Work on pages 7 and 9

REVISED WORK

great big kiss! I almost had a heart ~~(If this doesn't stop soon! Justin~~ ~~(found out I wasn't)~~ attack!

After mom left, I was searching around for something to do. When I opened my bottom drawer, I found that Justin had left a chocolate bar for me. I guess he felt sorry that I wasn't allowed to eat tonight. I know he can feel remorse but such ~~(sy)~~ (sic) sympathy! Justin giving up a chocolate bar is like Alexis Colby giving up designer clothes. I wonder what he's up to? I sincerely doubt that he's suddenly become nice. Justin has too much fun being bad. Perhaps he's sifted my room so much that he knows he's going to have to store up brownie points for my return home.

I really am being hard on him.

There is additional space for Revised Work on page 9.

REVISED WORK

He could have found out that he misses me. [illegible] (Oh-lova. Whatever his reasons, I think this is the sweetest thing he's [illegible] ever done. Tomorrow I give him back his chocolate bar but I'll always treasure the wrapper.

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

As Alberta's economy continue to sag many women are returning to work. Our city of Nalwen is no exception. We are not, however, helping these financially troubled people as we should.

Nalwen has a deplorable lack of good day care centres. It is hard enough for a reluctant mother to leave her children without worrying over the quality of their babysitter. The two day-cares that Nalwen does have are running far over their intended child limits. The waits to get children into these centers is up to two years.

There is additional space for Revised Work on pages 15 and 17.

REVISED WORK

I believe taking care of these children should be city councils top priority. Instead of spending money on new offices and buildings for city council ~~[illegible]~~ they should be allocating money to day care Centres. The jobs this would create would again cycle more money and hope into the community.

Sincerely,
Pete Jones.

There is additional space for Revised Work on page 17.

Section III: Response to Visual Communication – Question 1

1. Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.

REVISED WORK

One little girl stands up in front of the class to show her feelings about the music being played. This is evident from the cartoon by noting the record player (in the composition) and the arrangement of the other boys and girls. The little girl, unaware of the (classes) class' reaction, is very proud of herself. She smile(s) broadly as she demonstrates her interpretation skills.

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

Even from an early age we all have different ideas. The main idea of the cartoon is expressed by the pictures quite effectively. The little girl before the class is attempting to communicate the [illegible] image of a flower. Each of the other class members perceive a different picture from the flower. The other class members also have a different picture from each of the others. These different pictures illustrate the different ideas each child can have.

REVISED WORK

Images

Kevin has always been an overly popular, outgoing guy. Everyone seems to look up to him. Everything he does seems "right-on."

Kevin and I have gone out off and on for the past four years. Until now, I never got to know him so well. We've been seeing each other exclusively, which means we are together most of the time, in school and out.

This past month I have learned so much about Kevin. Even though he has a type of image that he thinks he has to keep up, he is really quite understanding and tender-hearted.

I never thought little things meant much to him, but I was definitely wrong. I never knew this until his hockey accident. — It was near the end of the season, not quite into the finals, when Kevin was facing off against a big, tough guy. The whistle blew, the puck was dropped and the two determined bodies smashed into each other at tremendous impact. Both were knocked down, but Kevin was down longer. I was sitting in the bleachers, beside myself.

here is additional space for Revised Work on pages 7 and 9.

REVISED WORK

I started to panic because when Kevin is knocked down on ice, he must have been hit very, very hard. My mind was blank and I didn't know what could be wrong. The other guys were standing around, along with the coaches, and referees. At this point, the assistant coach started skating towards the bench that I was sitting at. He called my name and asked me to come out onto the ice because Kevin is asking for me. Right then my heart started beating crazily. When I finally got down onto the ice, not caring what other girls were to say or anything, I went over towards the area where Kevin ~~he~~ was sprawled out. Kevin had his one arm around his chest and the other one was released outwards from his body. He looked up to me and said, "Michelle, take my hand, I need you now."

I dropped to the ice and all I could do was cry. Just to see him laying there, helpless made me upset. The men carried a stretcher onto the ice and carefully move him onto that.

The last thing Kevin said to me was, "Take

There is additional space for Revised Work on page 9.

my car to the ~~emergi~~ hospital and I hope to see you there."

This more less shocked me, because he wouldn't have never let me drive his car before, but in this case, I saw the difference

When I got to the hospital, I sat with the coach and a few guys from the team, patiently waiting to hear about Kevin's condition.

Two hours later, a tall, perky-looking doctor came into the waiting room and said, "That young man was hurt and hurt good, but he's going be alright. He broke six ribs, along with an almost ruptured spleen. No more hockey this year for him. He says "hi" to you all, but before he visits, he wants to see his girl."

I stepped forward and said. "I think that's me." I followed the doctor to the room Kevin was in. He told me not to be too long because Kevin needed a little rest before he could go anywhere. I look down at Kevin and before I could say anything, he exclaimed, "It's no big deal."

I said, "Kevin, I was so ~~stok~~ shocked that in that time of pain you experienced, you actualy asked me to come around. Didn't you think people were going to think you are too much in love or something?"

Kevin said, "When it happened all I could think of was you. I thought if something drastic was wrong to ruin my hockey future, you were there for me; I guess I needed the reassurance."

"I thought you would be the last person in the world you needs reassurance."

All he said was, "Expect the unexpected. Kevins behavior will always remain memorable to me because I never knew that side of him before. I like it!

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

June 13, 1986

Dear Sir:

I am writing this letter in concern of the need to renovate the community hall of Nalwen.

I am sure you will agree with me that this building should be restored to good condition.

If this hall was to be new again the people in our community would definitely get together more. It would bring us all a little closer. More activities would go on if we had our community hall in order.

I hope you will consider doing something about this problem. Also, I hope to hear your reply, if possible.

Yours sincerely,
Pat Jones
Pat Jones

There is additional space for Revised Work on pages 15 and 17.

Section III: Response to Visual Communication – Question 1

1. **Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.**

REVISED WORK

I choose the dark-haired girl who is at the front of the class. This determined girl is trying to look like a flower. What she dosen't know is that her fellow classmates are all thinking something else.

It seems like she thinks she's doing such a good job, just by the proud look on her face.

Like any other student, she's probably a little nervous having to be up in front of the other kids.

Section III: Response to Visual Communication – Question 2

2. **What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.**

REVISED WORK

The main idea being communicated by the cartoon is that everyone thinks different from each other. Like in this cartoon, everyone has a different idea of what the girl is trying to act out as a flower.

APPENDIX C. GRADE 12 DIPLOMA EXAMINATION - ENGLISH 33 PART A: WRITTEN EXPRESSION, JANUARY, AND JUNE, 1986

FORTIS ET LIBER

GRADE 12
DIPLOMA EXAMINATION

English 33

Part A: Written Response

January 1986

Alberta
EDUCATION

GRADE 12 DIPLOMA EXAMINATION
ENGLISH 33

PART A: Written Response

GENERAL INSTRUCTIONS

This examination consists of THREE sections. Read the WHOLE examination before you begin to write. Complete ALL sections.

TOTAL TIME: 2½ hours

Budget your time carefully.

The three sections of the test are as follows:

		Page Number
Section I:	Personal Response to Literature Suggested time: 75 minutes (1¼ hours) Value: 50% of this examination	2
Section II:	Functional Writing Suggested time: 45 minutes (¾ hour) Value: 30% of this examination	11
Section III:	Response to Visual Communication Suggested time: 30 minutes (½ hour) Value: 20% of this examination	18

You may use a DICTIONARY and a THESAURUS.

Space is provided for PLANNING AND DRAFTING and for REVISED WORK.

Please write your revised work in blue or black ink.

DO NOT WRITE YOUR NAME ANYWHERE IN THE TEST BOOKLET

JANUARY 1986

SECTION I: PERSONAL RESPONSE TO LITERATURE

Read the excerpt below and complete the assignment that follows.

from THE KITE

The novel, The Kite, *focuses on Daddy Sherry, a man approaching his 111th birthday and a journalist, David Lang, who has been sent out for an extended time to do a feature article on him. The excerpt that follows is from the final pages of the novel. It is Daddy's birthday. Keith, a young boy, has just given Daddy a kite, and they have just launched the kite on its first flight.*

Keith turned and stood with eyes uplifted to the kite hanging almost straight over him. Even as he walked back to Daddy and David it gave no indication of sinking.

"Gimme your arm."

David helped Daddy as he lowered himself to the ground, leaned his back against the great rounded side of the rock imbedded in the earth there.

"Here you are, Daddy." Keith held out the stick to the old man.

For several moments the three sat silently on the grass, staring up to the kite. "Let out more string, Daddy," Keith said. "I only unwound a couple hundred feet of it."

Daddy released the pressure on his thumbs and the stick began to twirl. The kite sagged, began to fall. Daddy clamped down his thumbs; the kite took heart, soared upwards once more. Alternately the old man held and released the string thinning from the stick. Before half the string was out the kite had found the higher, stronger wind so that Daddy could unwind without stopping, the kite climbing persistently with no altitude loss whatever, yearning ever upwards, shrinking with distance till finally the stick was bare, the kite a high stamp pasted against the cloudless sky.

"That there," Daddy said, "outa all the birthdays I ever had — an' outa all the presents I ever got on 'em — is the nicest one of all. Thanks."

"Okay," Keith said.

"Look at her up there — hangin' steady — pullin' real strong on this string. . . ."

"I made it myself," Keith said, "mainly — Mr. Lang showed me how."

"I know — I know. She's a nice balanced kite — steady." He gave a long pull at the string. "See that —"

"What?" Keith said.

"Never even dodged." He pulled on the string again. "Strong — she's a strong one. Before we put her up again we got to take at least a foot off of that tail — then she'll be strong an' she'll be steady an' she'll be acrobatic too." He pulled on the string. "When I do that she oughta loop the loop an' she don't. Aaaaaah — she's a lovely kite — maybe not even a foot off — half a foot might do it. . . ."

W.O. Mitchell

Everyone has given and received a variety of gifts in his or her lifetime. Gifts may be purchased or handmade by the giver; they may be expensive or inexpensive, practical or frivolous. Whatever their nature, however, gifts such as the kite given to Daddy Sherry are evidence of the fact that one person cares for another.

WRITE ABOUT AN INCIDENT THAT FOCUSES ON THE GIVING OR RECEIVING OF A SPECIAL GIFT. THE INCIDENT YOU CHOOSE MAY BE FROM YOUR OWN EXPERIENCE OR FROM THE EXPERIENCE OF SOMEONE YOU KNOW OR HAVE READ ABOUT. DO NOT USE THE INCIDENT THAT YOU HAVE JUST READ FROM *THE KITE*.

In your writing BE SURE to

- **describe the incident**
- **examine the feelings and attitudes created by the incident**
- **explain how this incident has affected you**

Guidelines for Writing

Present your ideas in a PROSE FORM that will make your writing interesting. For example, you might wish to present your ideas in the form of a letter, a journal entry, or a conversation.

To develop your ideas you might wish to use description, definition, reason, examples, or any combination of these and other suitable methods.

Read the hypothetical situation described below and the assignment that follows.

THE SITUATION

Your school is organizing a "Careers Day." You have been given the task of inviting a speaker to your school to represent the career or occupation that most interests you. Posters such as the one below have been put up at various places throughout your school.

> *Plan **NOW** For Your Future Occupation!*
>
> **CAREERS DAY**
>
> June 21, 1986
> 9:00 A.M. - 4:00 P.M.
>
> *Nalwen Composite High School*
> MAIN GYMNASIUM
>
> *LECTURES*
> *DEMONSTRATIONS*
> *WORKSHOPS*
> *PAMPHLETS*
>
> *DON'T MISS THIS IMPORTANT EVENT*

THE ASSIGNMENT

IN THE SPACE PROVIDED, WRITE A LETTER INVITING YOUR SPEAKER TO NALWEN COMPOSITE HIGH SCHOOL.

BE SURE THAT YOUR LETTER

- **demonstrates your interest in the career or occupation that your speaker will represent**
- **includes all of the information that your speaker will need to make his or her participation successful**

Use an appropriate tone in your writing.

PLEASE NOTE: Correct letter format has been provided beginning on page 13. Please give your speaker a name. SIGN YOUR LETTER "Pat Jones." DO NOT USE THE NAME OF YOUR OWN SCHOOL ANYWHERE IN YOUR WRITING.

Nalwen Composite High School
P.O. Box 203, Nalwen, Alberta. T5J 2R2

M ______________________________
16 Riverview Crescent
Nalwen, Alberta
T5J 2R2

January 13, 1986

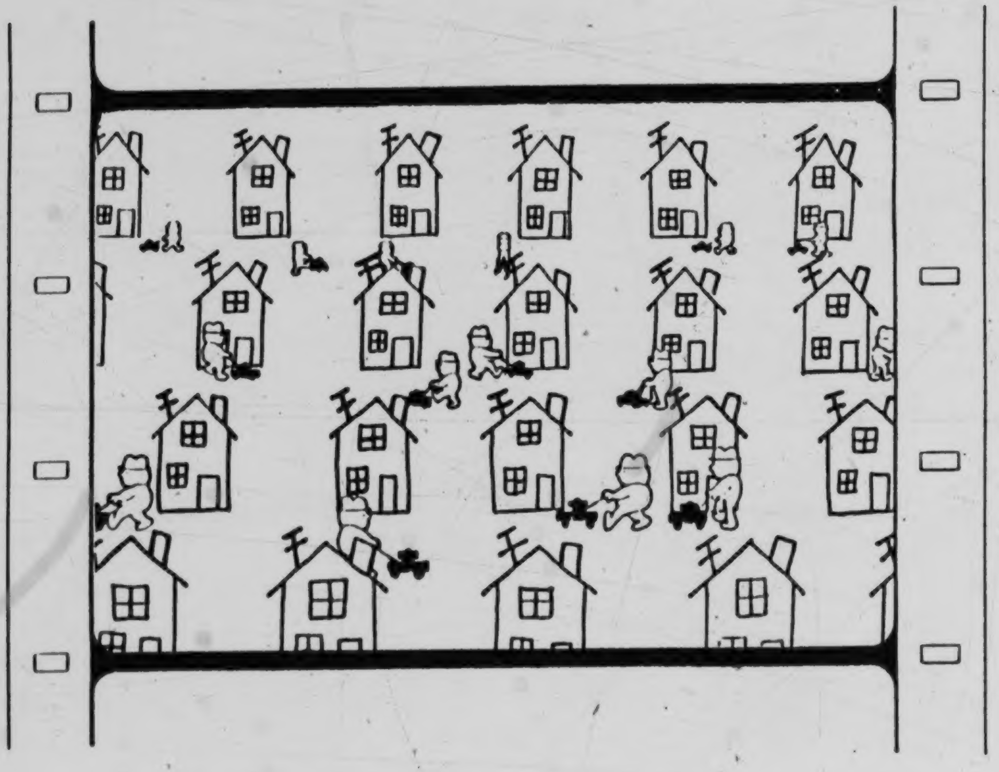
Dear M ___________________________:

There is additional space for Revised Work on pages 15 and 17.

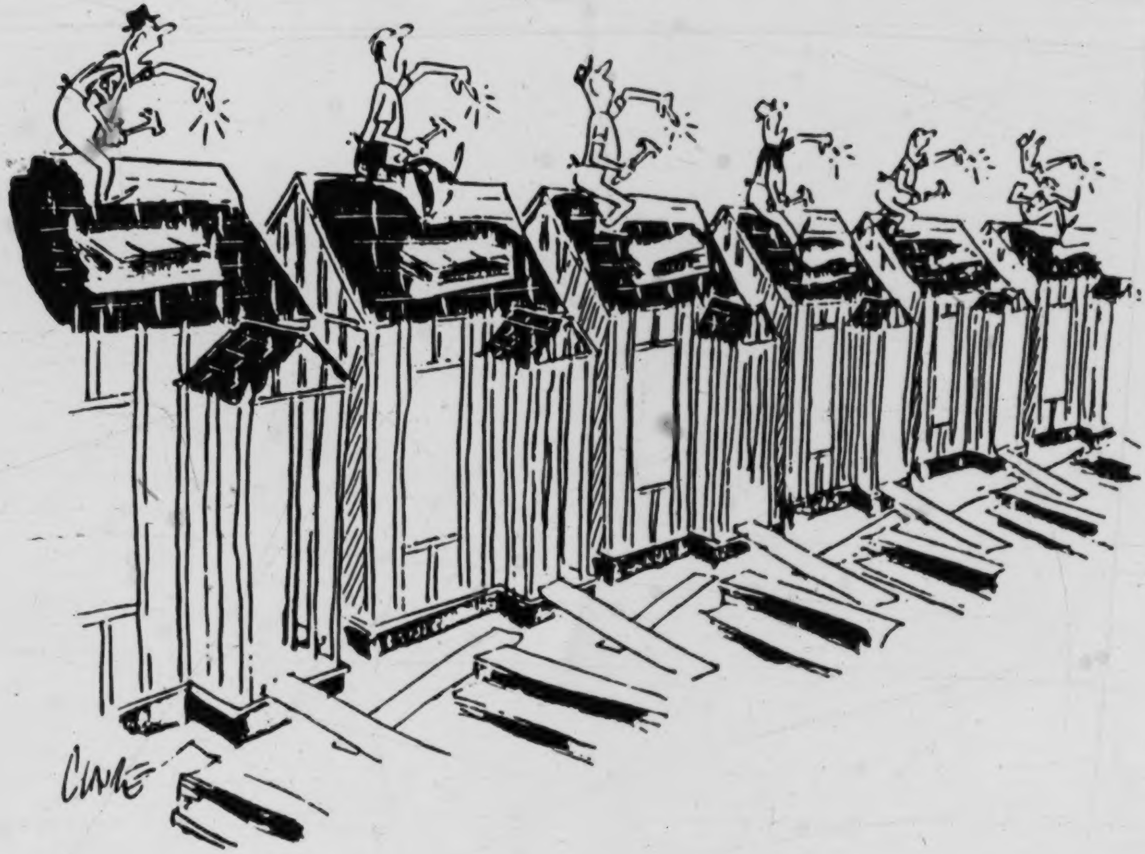
SECTION III: RESPONSE TO VISUAL COMMUNICATION

Examine the two selections that follow. Selection One is a single frame from the National Film Board of Canada's short animated film *The House That Jack Built*. Selection Two is a cartoon from *The New Yorker* magazine. Answer Question 1 on page 21 and Question 2 on page 23 in paragraph form.

Selection One:



Selection Two:



Section III: Response to Visual Communication – Question 1

1. Describe how you would feel if you were one of the characters in Selection One OR one of the characters in Selection Two. Support your answer with reasons that are consistent with the situation that is presented. Answer in paragraph form.

REVISED WORK

2. What is a main idea that is common to both Selection One AND Selection Two? Use details from the two selections to support your answer. Answer in paragraph form.

REVISED WORK

CREDITS

The Kite by W.O. Mitchell. Reprinted by permission of Macmillan of Canada, a Division of Canada Publishing Corporation.

From The National Film Board's short animation film *The House that Jack Built.*

Drawing by Claude; © 1960 The New Yorker Magazine, Inc.

FOR DEPARTMENT USE ONLY

M1 ☐

M2 ☐

M3 ☐

FOR DEPARTMENT USE ONLY

ENGLISH 33: PART A

ALBERTA EDUCATION STUDENT ID ☐☐☐☐☐☐☐☐☐

NAME: (LAST NAME) ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ (FIRST NAME) ☐☐☐☐☐☐☐☐☐☐☐☐ DATE OF BIRTH: Y ☐☐ M ☐☐ D ☐☐ SEX: ☐

PERMANENT MAILING ADDRESS: ____________________
(Apt./Street/Ave./P.O. Box) (Village/Town/City) (Postal Code)

SCHOOL CODE: ☐☐☐☐ SCHOOL: ____________________ SIGNATURE: ____________________

FOR DEPARTMENT USE ONLY

ENGLISH 33: PART A

FORTIS ET LIBER

GRADE 12
DIPLOMA EXAMINATION

English 33

Part A: Written Response

June 1986

Alberta
EDUCATION

GRADE 12 DIPLOMA EXAMINATION
ENGLISH 33

PART A: Written Response

GENERAL INSTRUCTIONS

This examination consists of THREE sections. Read the WHOLE examination before you begin to write. Complete ALL sections.

Total time: 2½ hours

Budget your time carefully.

The three sections of the test are as follows:

		Page Number
Section I:	Personal Response to Literature Suggested time: 75 minutes (1¼ hours) Value: 50% of this examination	2
Section II:	Functional Writing Suggested time: 45 minutes (¾ hour) Value: 30% of this examination	11
Section III:	Response to Visual Communication Suggested time: 30 minutes (½ hour) Value: 20% of this examination	19

You may use a DICTIONARY and a THESAURUS.

Space is provided for PLANNING AND DRAFTING and for REVISED WORK.

Please write your revised work in blue or black ink.

DO NOT WRITE YOUR NAME ANYWHERE IN THE TEST BOOKLET

JUNE 1986

SECTION I: PERSONAL RESPONSE TO LITERATURE

Read the poem below and complete the assignment that follows.

BRIAN SWANSON

One clean-sponge May
I cut a dead, damp cocoon
from a soft redwood fence
for Bad Brian Swanson,
who once beat up my brother
and threw a baseball
through Carlson's picture window.

Every time he sees me, Brian boasts cockily
that his parents are still paying for
the broken window; but, softer, almost shyly,
he always whispers that
the cocoon just hatched yesterday
and turned into a beautiful moth
with purple wings that shine in the dark.

David Thompson

Section I: Personal Response to Literature Assignment

There is often a difference between a person's public reputation and his or her private self. Occasionally a person will behave in a way that is different from what others would expect.

WRITE ABOUT A PERSON WHO, LIKE BRIAN SWANSON, HAS REVEALED AN UNEXPECTED SIDE OF HIMSELF OR HERSELF. YOU MAY WRITE ABOUT YOURSELF OR ABOUT SOMEONE YOU KNOW OR HAVE READ ABOUT. DO NOT WRITE ABOUT BRIAN SWANSON.

In your writing BE SURE to

- **describe the person's public reputation**
- **relate an incident that reveals an unexpected side of the person**
- **examine feelings and attitudes resulting from the person's unexpected behavior**
- **explain why the person's unexpected behavior is memorable to you**

Section I: Personal Response to Literature

Guidelines for Writing

You may present your ideas in any PROSE form that will make your writing interesting. For example, you might present your ideas in the form of a letter, a journal entry, or a conversation.

SECTION II: FUNCTIONAL WRITING

Read the following imaginary situation and complete the assignment that follows.

THE SITUATION

In a recent editorial, the editors of your local newspaper, *The Nalwen News*, have listed some community problems that they feel require immediate action by the community. Included in the list are

- a lack of recreational facilities for youth
- a lack of community services for senior citizens
- an absence of day-care centres
- a need for park development
- a need for ongoing snowplowing and road maintenance
- a need to upgrade garbage collection and disposal
- a need to renovate the community hall

The editors have asked each newspaper reader to write a letter that examines ONE of these problems or any other community problem that is of concern to the reader.

THE ASSIGNMENT

IDENTIFY ONE PROBLEM THAT IS OF CONCERN IN YOUR COMMUNITY. IN THE SPACE PROVIDED, WRITE A LETTER TO THE EDITORS OF *THE NALWEN NEWS* THAT ILLUSTRATES THE PROBLEM AS YOU HAVE EXPERIENCED IT.

In your letter

- **describe, in detail, the problem that is of concern to you**
- **tell what should be done about the problem you have described**

Use an appropriate tone in your writing.

PLEASE NOTE: Correct letter format has been provided beginning on page 13.

DO NOT NAME YOUR ACTUAL COMMUNITY OR DISTRICT. Use "Nalwen" as the name of your community or district.

DO NOT USE YOUR OWN NAME IN YOUR WRITING. Sign your letter "Pat Jones".

Section II: Functional Writing

P.O. Box 203
Nalwen, Alberta
T5J 2R2

The Editors
The Nalwen News
603 Block Terrace
Nalwen, Alberta
T5J 2R2

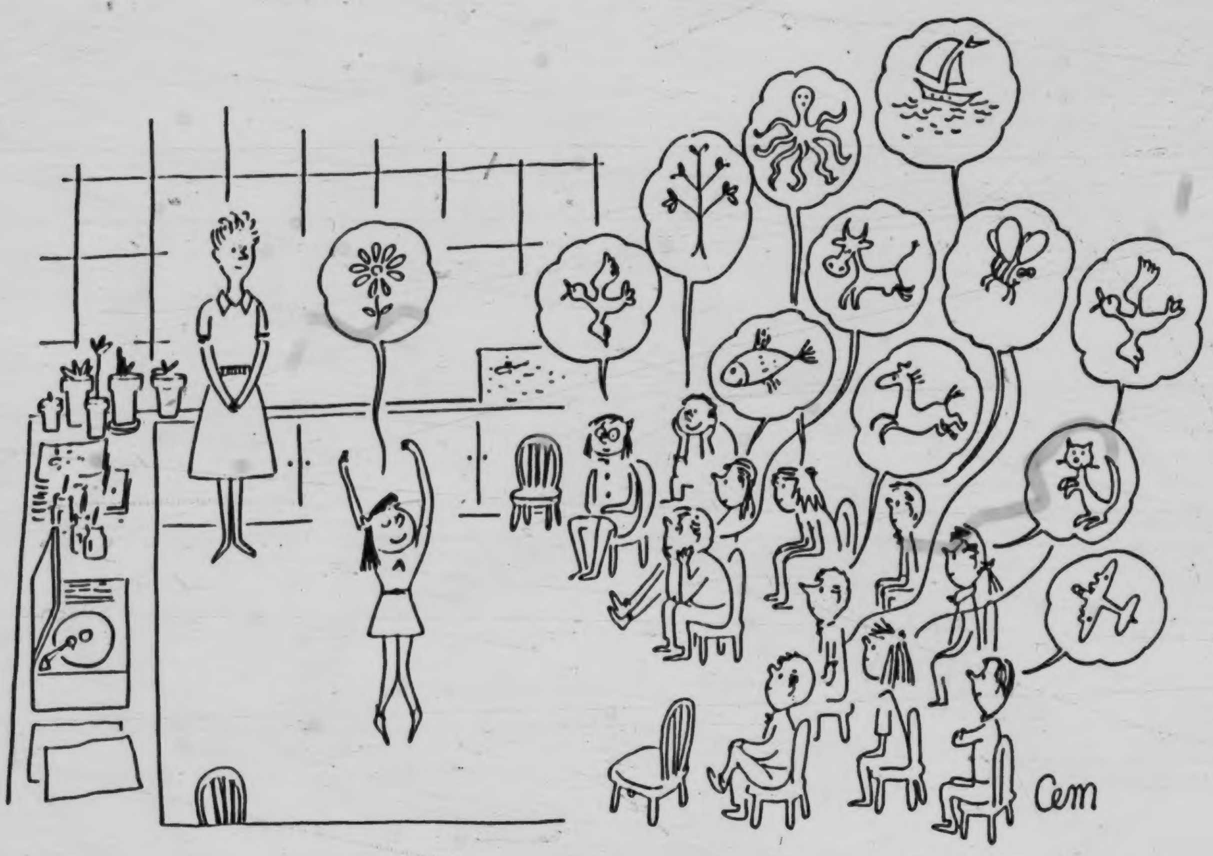
June 13, 1986

Dear Sir:

There is additional space for Revised Work on pages 15 and 17.

SECTION III: RESPONSE TO VISUAL COMMUNICATION

Examine the cartoon and answer the TWO questions that follow. Answer each question in paragraph form.



/305

Section III: Response to Visual Communication – Question 1

1. Choose ONE of the characters shown in the cartoon and describe what you think the character is feeling in this situation. Use specific details from the cartoon to support your ideas. Answer in paragraph form.

REVISED WORK

Section III: Response to Visual Communication – Question 2

2. What is a main idea being communicated by the cartoon? Use details from the cartoon to support your answer. Answer in paragraph form.

REVISED WORK

CREDITS

David Thompson. "Brian Swanson" from *Literary Cavalcade* (New York: Scholastic Magazines, Inc.) Volume 20 Number 8 May 1968. Copyright © 1968 by Scholastic Inc. Used with permission of the Scholastic Writing Awards Program.

Drawing by C.E.M.; © 1961 The New Yorker Magazine, Inc.

FOR DEPARTMENT USE ONLY

M1 ☐

M2 ☐

M3 ☐

FOR DEPARTMENT USE ONLY

ENGLISH 33: PART A

ALBERTA EDUCATION STUDENT ID ☐☐☐☐☐☐☐☐☐

NAME: (LAST NAME) ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐ (FIRST NAME) ☐☐☐☐☐☐☐☐☐☐☐☐ DATE OF BIRTH: Y ☐☐ M ☐☐ D ☐☐ SEX: ☐

PERMANENT MAILING ADDRESS: ______________________________
(Apt./Street/Ave./P.O. Box) (Village/Town/City) (Postal Code)

SCHOOL CODE: ☐☐☐☐ SCHOOL: ____________________ SIGNATURE: ____________________

FOR DEPARTMENT USE ONLY

ENGLISH 33: PART A

APPENDIX D.

Table D-1 MEAN CORRELATIONS BY WRITING TASK FOR ALL READERS (JANUARY, 1986)

Table D-2 MEAN CORRELATIONS BY WRITING TASK FOR ALL READERS (JUNE, 1986)

Table D-1

MEAN CORRELATIONS BY WRITING TASK FOR ALL READERS

JANUARY, 1986 (N = 75)

Reader #	Mean Corr. Section I	Mean Corr. Section II	Mean Corr. Section III
1	.769	.775	.884
2	.681	.763	.841
3	.662	.702	.845
4	.769	.798	.880
5	.652	.622	.858
6	.623	.495	.497
7	.793	.691	.852
8	.737	.313	.535
9	.720	.722	.853
10	.618	.738	.756
11	.839	.809	.892
12	.515	.538	.762
13	.794	.685	.893
14	.668	.732	.786
15	.663	.677	.872
16	.632	.802	.863
17	.612	.606	.840
18	.531	.806	.813
19	.666	.718	.780
20	.811	.604	.903
21	.819	.545	.769
22	.734	.744	.819
23	.783	.774	.889
24	.792	.764	.871
25	.710	.787	.810
26	.853	.782	.861
27	.820	.763	.903
28	.792	.343	.818
29	.671	.732	.828
30	.797	.780	.775
31	.639	.701	.849
32	.790	.767	.855
33	.802	.766	.838
34	.682	.807	.909
35	.766	.429	.834
36	.777	.699	.765
37	.844	.671	.851
38	.606	.668	.831
39	.727	.652	.789
40	.784	.550	.855
41	.715	.763	.891
42	.705	.462	.778
43	.847	.686	.890

Table D-1 (Cont'd)

Reader #	Mean Corr. Section I	Mean Corr. Section II	Mean Corr. Section III
44	.720	.775	.878
45	.820	.718	.886
46	.779	.549	.880
47	.804	.528	.804
48	.769	.743	.896
49	.802	.736	.820
50	.763	.542	.813
51	.766	.601	.791
52	.726	.492	.724
53	.800	.669	.841
54	.778	.559	.692
55	.784	.619	.841
56	.782	.635	.817
57	.773	.669	.847
58	.723	.718	.888
59	.759	.584	.817
60	.759	.689	.810
61	.548	.699	.881
62	.752	.790	.872
63	.706	.608	.881
64	.851	.612	.912
65	.770	.766	.893
66	.809	.612	.887
67	.714	.694	.900
68	.761	.684	.929
69	.790	.757	.872
70	.541	.559	.845
71	.766	.413	.820
72	.842	.787	.935
73	.840	.799	.892
74	.796	.682	.858
75	.685	.711	.863
$\overline{X}$ =	.740	.672	.835

Table D-2

MEAN CORRELATIONS BY WRITING TASK FOR ALL READERS

JUNE, 1986 (N = 64)

Reader #	Mean Corr. Section I	Mean Corr. Section II	Mean Corr. Section III
1	.615	.472	.716
2	.766	.562	.713
3	.771	.645	.582
4	.774	.680	.708
5	.708	.695	.714
6	.643	.669	.691
7	.741	.623	.784
8	.747	.752	.760
9	.760	.654	.726
10	.783	.737	.592
11	.705	.479	.606
12	.729	.614	.722
13	.798	.618	.761
14	.766	.634	.805
15	.514	.690	.491
16	.592	.633	.665
17	.581	.610	.774
18	.657	.571	.581
19	.597	.403	.650
20	.514	.201	.595
21	.782	.726	.688
22	.676	.715	.763
23	.717	.682	.760
24	.767	.653	.671
25	.682	.728	.725
26	.741	.570	.738
27	.758	.698	.716
28	.774	.556	.682
29	.678	.484	.733
30	.328	.651	.681
31	.701	.607	.745
32	.758	.510	.718
33	.728	.670	.711
34	.690	.733	.760
35	.790	.690	.810
36	.733	.555	.708
37	.775	.733	.780
38	.635	.557	.611
39	.562	.518	.756
40	.640	.673	.628
41	.743	.484	.721
42	.720	.693	.650
43	.473	.440	.695

Table D-2 (Cont'd)

Reader #	Mean Corr. Section I	Mean Corr. Section II	Mean Corr. Section III
44	.804	.523	.660
45	.800	.567	.771
46	.787	.635	.635
47	.569	.337	.733
48	.774	.607	.719
49	.414	.359	.476
50	.675	.498	.725
51	.680	.606	.611
52	.639	.628	.649
53	.772	.655	.636
54	.522	.543	.787
55	.770	.574	.781
56	.744	.611	.735
57	.754	.621	.617
58	.716	.591	.716
59	.750	.374	.763
60	.803	.744	.807
61	.710	.564	.610
62	.678	.337	.642
63	.447	.544	.662
64	.702	.644	.643
$\overline{X}$ =	.689	.591	.695

VITA

NAME: Adell Mona Nyberg

PLACE OF BIRTH: Edmonton, Alberta, Canada

YEAR OF BIRTH: 1952

POST-SECONDARY EDUCATION:

The University of Alberta
Edmonton, Alberta
1970 - 1974 B.Ed. with distinction

The University of Alberta
Edmonton, Alberta
1976 - 1977 M.Ed.

HONORS AND AWARDS:

Province of Alberta Scholarship
1972, 1973

First Class Standing, The University of Alberta
1972, 1973, 1974

Social Sciences and Humanities Research Council of Canada
Doctoral Fellowship
1986, 1987

Graduate Faculty Fellowship, The University of Alberta
1986

RELATED WORK EXPERIENCE

English Teacher
County of Parkland No. 31
1974 - 1976, 1977 - 1982, 1983 - 1985

Practicum Associate
Department of Field Services
The University of Alberta
1982 - 1983

English Department Head
County of Parkland No. 31
1984 - 1985

Sessional Lecturer
Department of Secondary Education
The University of Alberta
1985 - 1986

Diploma Examination Group Leader
Student Evaluation Branch
Alberta Education
January, June, and August, 1984 - 1987

PUBLICATIONS

Nyberg, Verner R., and Adell M. Nyberg. *Alberta Essay Scales:* Models. Edmonton: Alberta Education, 1977.

Nyberg, Verner R., and Adell M. Nyberg. *Technical Report on Alberta Essay Scales: Models*. Edmonton: Alberta Education, 1977.

Nyberg, Adell, Patti Penner, and Robin Smith. *The Student Survival Handbook*. Stony Plain, Alberta: The County of Parkland No. 31, 1978.

Nyberg, Verner, and Adell Nyberg. "Reliability of Alberta Essay Scales." *Alberta Journal of Educational Research*, Vol XXVI, No. 1 (March, 1980), 64-67.

Nyberg, Adell, Patti Penner, and Robin Smith. *The Student Survival Handbook*. 2nd. ed. Stony Plain, Alberta: The County of Parkland No. 31, 1981.

Nyberg, Adell. "Measuring Essay Grading Standards Over Time." *Alberta English '81*, Vol. 20, No. 2 (Fall, 1981), 25-26.

Nyberg, Verner R., and Adell M. Nyberg. *Alberta Essay Scales: Models*. Edmonton: Faculty of Education, U. of Alberta, 1982.

Nyberg, Verner R., and Adell M. Nyberg. *Technical Report on Alberta Essay Scales: Models*. Edmonton: Faculty of Education, U. of Alberta, 1982.

Nyberg, Adell. "Alberta High School Language Arts Curriculum." *The English Quarterly*, Vol. XV, No. 4 (Winter 1982-83), 46-50.

Nyberg, Adell. "Now Hear This!" *Alberta English '85*, Vol. 23, No. 1 (Spring, 1985), 20-22.

Bell, James B., and Adell M. Nyberg. *Reflections, Revised Edition, Teacher's Manual*. Toronto: John Wiley and Sons, 1985.

Bell, James B., and Adell M. Nyberg. *Explorations, Revised Edition, Teacher's Manual*. Toronto: John Wiley and Sons, 1985.

Nyberg, Adell M. "Hear No Evil, See No Evil, Speak No Evil: Censorship and the Teaching of Literature." Paper delivered at the Third Occasional Conference of Western Canadian Language Arts Teacher Educators, April, 1986.

Assisted James B. Bell with the writing of the first and second editions of *The Little English Handbook for Canadians* in 1977 and 1982.